GENE PRODUCTS DIFFERENTIALLY EXPRESSED IN CANCEROUS CELLS AND THEIR METHODS OF USE II

SEQUENCE LISTING AND TABLES

A Sequence Listing is provided as part of this specification on triplicate compact discs, filed concurrently herewith, which compact discs named "Copy 1", "Copy 2", and "CRF" each of which compact discs contain the following file: "SEQLIST.TXT", created February 10, 2004, of 18 Megabytes, which is incorporated herein by reference in its entirety.

The present application also incorporates by reference Tables 2, 17, 18, 41A, 41B, 70A, 70B, 83, 84, 85, 86, 106, 107A, 107B, 110, 114, 130, 131A, 131B, 133, 134, 141, 143, 151 and 162 contained on duplicate compact discs filed concurrently herewith, which compact discs are labeled "Atty Docket 21302.001 Tables Copy 1" and "Atty Docket 21302.001 Tables Copy 2". The details of these Tables are further described later in this disclosure. These compact discs were created on February 10, 2004. The sizes of the Tables are as follows: Table 2: 147 kilobytes; Table 17: 344 kilobytes; Table 18: 372 kilobytes; Table 41A: 98 kilobytes; Table 41B: 41 kilobytes; Table 70A: 90 kilobytes; Table 70B: 72 kilobytes; Table 83: 60 kilobytes; Table 84: 94 kilobytes; Table 85: 251 kilobytes; Table 86: 232 kilobytes; Table 106: 148 kilobytes; Table 107A: 193 kilobytes; Table 107B: 138 kilobytes; Table 110: 278 kilobytes; Table 114: 11 kilobytes; Table 130: 395 kilobytes; Table 131A: 569 kilobytes; Table 131B: 354 kilobytes; Table 133: 40 kilobytes; Table 134: 8 kilobytes; Table 141: 402 kilobytes; Table 143: 98 kilobytes; Table 151: 8 kilobytes; and Table 162: 684 kilobytes.

FIELD OF THE INVENTION

The present invention relates to polynucleotides of human origin in substantially isolated form and gene products that are differentially expressed in cancer cells, and uses thereof.

BACKGROUND OF THE INVENTION

Cancer, like many diseases, is not the result of a single, well-defined cause, but rather can be viewed as several diseases, each caused by different aberrations in informational pathways, that ultimately result in apparently similar pathologic phenotypes. Identification of

polynucleotides that correspond to genes that are differentially expressed in cancerous, pre-cancerous, or low metastatic potential cells relative to normal cells of the same tissue type, provides the basis for diagnostic tools, facilitates drug discovery by providing for targets for candidate agents, and further serves to identify therapeutic targets for cancer therapies that are more tailored for the type of cancer to be treated.

Identification of differentially expressed gene products also furthers the understanding of the progression and nature of complex diseases such as cancer, and is key to identifying the genetic factors that are responsible for the phenotypes associated with development of, for example, the metastatic phenotype. Identification of gene products that are differentially expressed at various stages, and in various types of cancers, can both provide for early diagnostic tests, and further serve as therapeutic targets. Additionally, the product of a differentially expressed gene can be the basis for screening assays to identify chemotherapeutic agents that modulate its activity (e.g. its expression, biological activity, and the like).

Early disease diagnosis is of central importance to halting disease progression, and reducing morbidity. Analysis of a patient's tumor to identify the gene products that are differentially expressed, and administration of therapeutic agent(s) designed to modulate the activity of those differentially expressed gene products, provides the basis for more specific, rational cancer therapy that may result in diminished adverse side effects relative to conventional therapies. Furthermore, confirmation that a tumor poses less risk to the patient (*e.g.*, that the tumor is benign) can avoid unnecessary therapies. In short, identification of genes and the encoded gene products that are differentially expressed in cancerous cells can provide the basis of therapeutics, diagnostics, prognostics, therametrics, and the like.

For example, breast cancer is a leading cause of death among women. One of the priorities in breast cancer research is the discovery of new biochemical markers that can be used for diagnosis, prognosis and monitoring of breast cancer. The prognostic usefulness of these markers depends on the ability of the marker to distinguish between patients with breast cancer who require aggressive therapeutic treatment and patients who should be monitored.

While the pathogenesis of breast cancer is unclear, transformation of non-tumorigenic breast epithelium to a malignant phenotype may be the result of genetic factors, especially in women under 30 (Miki, et al., Science, 266: 66-71, 1994). However, it is likely that other, non-genetic factors are also significant in the etiology of the disease. Regardless of its origin, breast

cancer morbidity increases significantly if a lesion is not detected early in its progression. Thus, considerable effort has focused on the elucidation of early cellular events surrounding transformation in breast tissue. Such effort has led to the identification of several potential breast cancer markers.

Thus, the identification of new markers associated with cancer, for example, breast cancer, and the identification of genes involved in transforming cells into the cancerous phenotype, remains a significant goal in the management of this disease. In exemplary aspects, the invention described herein provides cancer diagnostics, prognostics, therametrics, and therapeutics based upon polynucleotides and/or their encoded gene products.

SUMMARY OF THE INVENTION

The present invention provides methods and compositions useful in detection of cancerous cells, identification of agents that modulate the phenotype of cancerous cells, and identification of therapeutic targets for chemotherapy of cancerous cells. Cancerous, breast, colon and prostate cells are of particular interest in each of these aspects of the invention. More specifically, the invention provides polynucleotides in substantially isolated form, as well as polypeptides encoded thereby, that are differentially expressed in cancer cells. Also provided are antibodies that specifically bind the encoded polypeptides. These polynucleotides, polypeptides and antibodies are thus useful in a variety of diagnostic, therapeutic, and drug discovery methods. In some embodiments, a polynucleotide that is differentially expressed in cancer cells can be used in diagnostic assays to detect cancer cells. In other embodiments, a polynucleotide that is differentially expressed in cancer cells, and/or a polypeptide encoded thereby, is itself a target for therapeutic intervention.

Accordingly, the invention features an isolated polynucleotide comprising a nucleotide sequence having at least 90% sequence identity to an identifying sequence of any one of the sequences set forth herein or a degenerate variant thereof. In related aspects, the invention features recombinant host cells and vectors comprising the polynucleotides of the invention, as well as isolated polypeptides encoded by the polynucleotides of the invention and antibodies that specifically bind such polypeptides.

In other aspects, the invention provides a method for detecting a cancerous cell. In general, the method involves contacting a test sample obtained from a cell that is suspected of

being a cancer cell with a probe for detecting a gene product differentially expressed in cancer. Many embodiments of the invention involve a gene identifiable by or comprising a sequence selected from the group consisting of SEQ ID NOS: 1-23767, contacting the probe and the gene product for a time sufficient for binding of the probe to the gene product; and comparing a level of binding of the probe to the sample with a level of probe binding to a control sample obtained from a control cell of known cancerous state. A modulated (i.e. increased or decreased) level of binding of the probe in the test cell sample relative to the level of binding in a control sample is indicative of the cancerous state of the test cell. In certain embodiments, the level of binding of the probe in the test cell sample, usually in relation to at least one control gene, is similar to binding of the probe to a cancerous cell sample. In certain other embodiments, the level of binding of the probe in the test cell sample, usually in relation to at least one control gene, is different, i.e. opposite, to binding of the probe to a non-cancerous cell sample. In specific embodiments, the probe is a polynucleotide probe and the gene product is nucleic acid. In other specific embodiments, the gene product is a polypeptide. In further embodiments, the gene product or the probe is immobilized on an array.

In another aspect, the invention provides a method for assessing the cancerous phenotype (*e.g.*, metastasis, metastatic potential, aberrant cellular proliferation, and the like) of a cell comprising detecting expression of a gene product in a test cell sample, wherein the gene comprises or is identifiable using a sequence selected from the group consisting of SEQ ID NOS: 1-23767; and comparing a level of expression of the gene product in the test cell sample with a level of expression of the gene in a control cell sample. Comparison of the level of expression of the gene in the test cell sample relative to the level of expression in the control cell sample is indicative of the cancerous phenotype of the test cell sample. In specific embodiments, detection of gene expression is by detecting a level of an RNA transcript in the test cell sample. In other specific embodiments detection of expression of the gene is by detecting a level of a polypeptide in a test sample.

In another aspect, the invention provides a method for suppressing or inhibiting a cancerous phenotype of a cancerous cell, the method comprising introducing into a mammalian cell an expression modulatory agent (e.g. an antisense molecule, small molecule, antibody, neutralizing antibody, inhibitory RNA molecule, etc.) to inhibit expression of a gene identified by a sequence selected from the group consisting of SEQ ID NOS: 1-23767. Inhibition of

expression of the gene inhibits development of a cancerous phenotype in the cell. In specific embodiments, the cancerous phenotype is metastasis, aberrant cellular proliferation relative to a normal cell, or loss of contact inhibition of cell growth. In the context of this invention "expression" of a gene is intended to encompass the expression of an activity of a gene product, and, as such, inhibiting expression of a gene includes inhibiting the activity of a product of the gene.

In another aspect, the invention provides a method for assessing the tumor burden of a subject, the method comprising detecting a level of a differentially expressed gene product in a test sample from a subject suspected of or having a tumor, the differentially expressed gene product identified by or comprising a sequence selected from the group consisting of SEQ ID NOS: 1-23767. Detection of the level of the gene product in the test sample is indicative of the tumor burden in the subject.

In another aspect, the invention provides a method for identifying agents that modulate (i.e. increase or decrease) the biological activity of a gene product differentially expressed in a cancerous cell, the method comprising contacting a candidate agent with a differentially expressed gene product, the differentially expressed gene product corresponding to a sequence selected from the group consisting of SEQ ID NOS: 1-23767; and detecting a modulation in a biological activity of the gene product relative to a level of biological activity of the gene product in the absence of the candidate agent. In specific embodiments, the detecting is by identifying an increase or decrease in expression of the differentially expressed gene product. In other specific embodiments, the gene product is mRNA or cDNA prepared from the mRNA gene product. In further embodiments, the gene product is a polypeptide.

In another aspect, the invention provides a method of inhibiting growth of a tumor cell by modulating expression of a gene product, where the gene product is encoded by a gene identified by a sequence selected from the group consisting of: SEQ ID NOS:1-23767.

These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A-1B is a comparison of SEQ ID NO:15951 and clone H72034 (SEQ ID NO:15983).

Figure 2 is a comparison of SEQ ID NO:15982 and clone AA707002 (SEQ ID NO:15984).

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides polynucleotides, as well as polypeptides encoded thereby, that are differentially expressed in cancer cells. Methods are provided in which these polynucleotides and polypeptides are used for detecting and reducing the growth of cancer cells. Also provided are methods in which the polynucleotides and polypeptides of the invention are used in a variety of diagnostic and therapeutic applications for cancer. The invention finds use in the prevention, treatment, detection or research into any cancer, including prostrate, pancreas, colon, brain, lung, breast, bone, skin cancers, etc.

Before the present invention is described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications and patent applications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes a plurality of such polynucleotides and reference to "the cancer cell" includes reference to one or more cells and equivalents thereof known to those skilled in the art, and so forth.

The publications and applications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

Definitions

The terms "polynucleotide" and "nucleic acid", used interchangeably herein, refer to polymeric forms of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, these terms include, but are not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. These terms further include, but are not limited to, mRNA or cDNA that comprise intronic sequences (see, *e.g.*, Niwa et al. (1999) Cell 99(7):691-702). The backbone of the polynucleotide can comprise sugars and phosphate groups (as may typically be found in RNA or DNA), or modified or substituted sugar or phosphate groups. Alternatively, the backbone of the polynucleotide can comprise a polymer of synthetic subunits such as phosphoramidites and thus can be an oligodeoxynucleoside phosphoramidate or a mixed phosphoramidate-phosphodiester oligomer. Peyrottes *et al.* (1996) *Nucl. Acids Res.* 24:1841-1848; Chaturvedi *et al.* (1996) *Nucl. Acids Res.* 24:2318-2323. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs, uracyl, other sugars, and linking groups such as fluororibose and thioate, and nucleotide branches. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications included in this definition are caps, substitution of one or more of the naturally occurring nucleotides with an analog, and introduction of means for attaching the polynucleotide to proteins, metal ions, labeling components, other polynucleotides, or a solid support. The term "polynucleotide" also encompasses peptidic nucleic acids (Pooga et al Curr Cancer Drug Targets. (2001) 1:231-9).

A "gene product" is a biopolymeric product that is expressed or produced by a gene. A gene product may be, for example, an unspliced RNA, an mRNA, a splice variant mRNA, a

polypeptide, a post-translationally modified polypeptide, a splice variant polypeptide etc. Also encompassed by this term is biopolymeric products that are made using an RNA gene product as a template (i.e. cDNA of the RNA). A gene product may be made enzymatically, recombinantly, chemically, or within a cell to which the gene is native. In many embodiments, if the gene product is proteinaceous, it exhibits a biological activity. In many embodiments, if the gene product is a nucleic acid, it can be translated into a proteinaceous gene product that exhibits a biological activity.

A composition (e.g. a polynucleotide, polypeptide, antibody, or host cell) that is "isolated" or "in substantially isolated form" refers to a composition that is in an environment different from that in which the composition naturally occurs. For example, a polynucleotide that is in substantially isolated form is outside of the host cell in which the polynucleotide naturally occurs, and could be a purified fragment of DNA, could be part of a heterologous vector, or could be contained within a host cell that is not a host cell from which the polynucleotide naturally occurs. The term "isolated" does not refer to a genomic or cDNA library, whole cell total protein or mRNA preparation, genomic DNA preparation, or an isolated human chromosome. A composition which is in substantially isolated form is usually substantially purified.

As used herein, the term "substantially purified" refers to a compound (*e.g.,* a polynucleotide, a polypeptide or an antibody, etc.,) that is removed from its natural environment and is usually at least 60% free, preferably 75% free, and most preferably 90% free from other components with which it is naturally associated. Thus, for example, a composition containing A is "substantially free of" B when at least 85% by weight of the total A+B in the composition is A. Preferably, A comprises at least about 90% by weight of the total of A+B in the composition, more preferably at least about 95% or even 99% by weight. In the case of polynucleotides, "A" and "B" may be two different genes positioned on different chromosomes or adjacently on the same chromosome, or two isolated cDNA species, for example.

The terms "polypeptide" and "protein", interchangeably used herein, refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. The term includes fusion proteins, including, but not limited to, fusion proteins with a heterologous amino acid sequence, fusions with heterologous and

homologous leader sequences, with or without N-terminal methionine residues; immunologically tagged proteins; and the like.

"Heterologous" refers to materials that are derived from different sources (*e.g.*, from different genes, different species, etc.).

As used herein, the terms “a gene that is differentially expressed in a cancer cell,” and “a polynucleotide that is differentially expressed in a cancer cell” are used interchangeably herein, and generally refer to a polynucleotide that represents or corresponds to a gene that is differentially expressed in a cancerous cell when compared with a cell of the same cell type that is not cancerous, e.g., mRNA is found at levels at least about 25%, at least about 50% to about 75%, at least about 90%, at least about 1.5-fold, at least about 2-fold, at least about 5-fold, at least about 10-fold, or at least about 50-fold or more, different (*e.g.*, higher or lower). The comparison can be made in tissue, for example, if one is using in situ hybridization or another assay method that allows some degree of discrimination among cell types in the tissue. The comparison may also or alternatively be made between cells removed from their tissue source.

"Differentially expressed polynucleotide" as used herein refers to a nucleic acid molecule (RNA or DNA) comprising a sequence that represents a differentially expressed gene, *e.g.*, the differentially expressed polynucleotide comprises a sequence (*e.g.*, an open reading frame encoding a gene product; a non-coding sequence) that uniquely identifies a differentially expressed gene so that detection of the differentially expressed polynucleotide in a sample is correlated with the presence of a differentially expressed gene in a sample. “Differentially expressed polynucleotides” is also meant to encompass fragments of the disclosed polynucleotides, *e.g.*, fragments retaining biological activity, as well as nucleic acids homologous, substantially similar, or substantially identical (*e.g.*, having about 90% sequence identity) to the disclosed polynucleotides.

"Corresponds to" or "represents" when used in the context of, for example, a polynucleotide or sequence that "corresponds to" or "represents" a gene means that at least a portion of a sequence of the polynucleotide is present in the gene or in the nucleic acid gene product (*e.g.*, mRNA or cDNA). A subject nucleic acid may also be “identified” by a polynucleotide if the polynucleotide corresponds to or represents the gene. Genes identified by a polynucleotide may have all or a portion of the identifying sequence wholly present within an exon of a genomic sequence of the gene, or different portions of the sequence of the

polynucleotide may be present in different exons (*e.g.*, such that the contiguous polynucleotide sequence is present in an mRNA, either pre- or post-splicing, that is an expression product of the gene). In some embodiments, the polynucleotide may represent or correspond to a gene that is modified in a cancerous cell relative to a normal cell. The gene in the cancerous cell may contain a deletion, insertion, substitution, or translocation relative to the polynucleotide and may have altered regulatory sequences, or may encode a splice variant gene product, for example. The gene in the cancerous cell may be modified by insertion of an endogenous retrovirus, a transposable element, or other naturally occurring or non-naturally occurring nucleic acid. In most cases, a polynucleotide corresponds to or represents a gene if the sequence of the polynucleotide is most identical to the sequence of a gene or its product (e.g. mRNA or cDNA) as compared to other genes or their products. In most embodiments, the most identical gene is determined using a sequence comparison of a polynucleotide to a database of polynucleotides (e.g. GenBank) using the BLAST program at default settings For example, if the most similar gene in the human genome to an exemplary polynucleotide is the protein kinase C gene, the exemplary polynucleotide corresponds to protein kinase C. In most cases, the sequence of a fragment of an exemplary polynucleotide is at least 95%, 96%, 97%, 98%, 99% or up to 100% identical to a sequence of at least 15, 20, 25, 30, 35, 40, 45, or 50 contiguous nucleotides of a corresponding gene or its product (mRNA or cDNA), when nucleotides that are "N" represent G, A, T or C.

An "identifying sequence" is a minimal fragment of a sequence of contiguous nucleotides that uniquely identifies or defines a polynucleotide sequence or its complement. In many embodiments, a fragment of a polynucleotide uniquely identifies or defines a polynucleotide sequence or its complement. In some embodiments, the entire contiguous sequence of a gene, cDNA, EST, or other provided sequence is an identifying sequence.

"Diagnosis" as used herein generally includes determination of a subject's susceptibility to a disease or disorder, determination as to whether a subject is presently affected by a disease or disorder, prognosis of a subject affected by a disease or disorder (*e.g.*, identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy), and use of therametrics (*e.g.*, monitoring a subject's condition to provide information as to the effect or efficacy of therapy).

As used herein, the term “a polypeptide associated with cancer” refers to a polypeptide encoded by a polynucleotide that is differentially expressed in a cancer cell.

The term “biological sample” encompasses a variety of sample types obtained from an organism and can be used in a diagnostic or monitoring assay. The term encompasses blood and other liquid samples of biological origin, solid tissue samples, such as a biopsy specimen or tissue cultures or cells derived therefrom and the progeny thereof. The term encompasses samples that have been manipulated in any way after their procurement, such as by treatment with reagents, solubilization, or enrichment for certain components. The term encompasses a clinical sample, and also includes cells in cell culture, cell supernatants, cell lysates, serum, plasma, biological fluids, and tissue samples.

The terms "treatment", "treating", "treat" and the like are used herein to generally refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete stabilization or cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease in a mammal, particularly a human, and includes: (a) preventing the disease or symptom from occurring in a subject which may be predisposed to the disease or symptom but has not yet been diagnosed as having it; (b) inhibiting the disease symptom, i.e., arresting its development; or (c) relieving the disease symptom, i.e., causing regression of the disease or symptom.

The terms “individual,” "subject," “host,” and "patient," used interchangeably herein and refer to any mammalian subject for whom diagnosis, treatment, or therapy is desired, particularly humans. Other subjects may include cattle, dogs, cats, guinea pigs, rabbits, rats, mice, horses, and the like.

A "host cell", as used herein, refers to a microorganism or a eukaryotic cell or cell line cultured as a unicellular entity which can be, or has been, used as a recipient for a recombinant vector or other transfer polynucleotides, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

The terms "cancer", "neoplasm", "tumor", and "carcinoma", are used interchangeably herein to refer to cells which exhibit relatively autonomous growth, so that they exhibit an

aberrant growth phenotype characterized by a significant loss of control of cell proliferation. In general, cells of interest for detection or treatment in the present application include precancerous (*e.g.*, benign), malignant, pre-metastatic, metastatic, and non-metastatic cells. Detection of cancerous cells is of particular interest.

The term "normal" as used in the context of "normal cell," is meant to refer to a cell of an untransformed phenotype or exhibiting a morphology of a non-transformed cell of the tissue type being examined.

"Cancerous phenotype" generally refers to any of a variety of biological phenomena that are characteristic of a cancerous cell, which phenomena can vary with the type of cancer. The cancerous phenotype is generally identified by abnormalities in, for example, cell growth or proliferation (*e.g.*, uncontrolled growth or proliferation), regulation of the cell cycle, cell mobility, cell-cell interaction, or metastasis, etc.

"Therapeutic target" generally refers to a gene or gene product that, upon modulation of its activity (*e.g.*, by modulation of expression, biological activity, and the like), can provide for modulation of the cancerous phenotype.

As used throughout, "modulation" is meant to refer to an increase or a decrease in the indicated phenomenon (*e.g.*, modulation of a biological activity refers to an increase in a biological activity or a decrease in a biological activity).

POLYNUCLEOTIDE COMPOSITIONS

The present invention provides isolated polynucleotides that contain nucleic acids that are differentially expressed in cancer cells. The polynucleotides, as well as any polypeptides encoded thereby, find use in a variety of therapeutic and diagnostic methods.

The scope of the invention with respect to compositions containing the isolated polynucleotides useful in the methods described herein includes, but is not necessarily limited to, polynucleotides having (i.e., comprising) a sequence set forth in any one of the polynucleotide sequences provided herein, or fragment thereof; polynucleotides obtained from the biological materials described herein or other biological sources (particularly human sources) by hybridization under stringent conditions (particularly conditions of high stringency); genes corresponding to the provided polynucleotides; cDNAs corresponding to the provided polynucleotides; variants of the provided polynucleotides and their corresponding genes,

particularly those variants that retain a biological activity of the encoded gene product (*e.g.*, a biological activity ascribed to a gene product corresponding to the provided polynucleotides as a result of the assignment of the gene product to a protein family(ies) and/or identification of a functional domain present in the gene product). Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here. "Polynucleotide" and "nucleic acid" as used herein with reference to nucleic acids of the composition is not intended to be limiting as to the length or structure of the nucleic acid unless specifically indicated.

The invention features polynucleotides that represent genes that are expressed in human tissue, specifically polynucleotides that are differentially expressed in tissues containing cancerous cells. Nucleic acid compositions described herein of particular interest are at least about 15 bp in length, at least about 30 bp in length, at least about 50 bp in length, at least about 100 bp, at least about 200 bp in length, at least about 300 bp in length, at least about 500 bp in length, at least about 800 bp in length, at least about 1 kb in length, at least about 2.0 kb in length, at least about 3.0 kb in length, at least about 5 kb in length, at least about 10 kb in length, at least about 50kb in length and are usually less than about 200 kb in length. These polynucleotides (or polynucleotide fragments) have uses that include, but are not limited to, diagnostic probes and primers as starting materials for probes and primers, as discussed herein.

The subject polynucleotides usually comprise a sequence set forth in any one of the polynucleotide sequences provided herein, for example, in the sequence listing, incorporated by reference in a table (e.g. by an NCBI accession number), a cDNA deposited at the A.T.C.C., or a fragment or variant thereof. A "fragment" or "portion" of a polynucleotide is a contiguous sequence of residues at least about 10 nt to about 12 nt, 15 nt, 16 nt, 18 nt or 20 nt in length, usually at least about 22 nt, 24 nt, 25 nt, 30 nt, 40 nt, 50 nt, 60nt, 70 nt, 80 nt, 90 nt, 100 nt to at least about 150 nt, 200 nt, 250 nt, 300 nt, 350 nt, 400 nt, 500 nt, 800 nt or up to about 1000 nt, 1500 or 2000 nt in length. In some embodiments, a fragment of a polynucleotide is the coding sequence of a polynucleotide. A fragment of a polynucleotide may start at position 1 (i.e. the first nucleotide) of a nucleotide sequence provided herein, or may start at about position 10, 20, 30, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1500 or 2000, or an ATG translational initiation codon of a nucleotide sequence provided herein. In this context "about" includes the particularly recited value or a value larger or smaller by several (5,

4, 3, 2, or 1) nucleotides. The described polynucleotides and fragments thereof find use as hybridization probes, PCR primers, BLAST probes, or as an identifying sequence, for example.

The subject nucleic acids may be variants or degenerate variants of a sequence provided herein. In general, a variants of a polynucleotide provided herein have a fragment of sequence identity that is greater than at least about 65%, greater than at least about 70%, greater than at least about 75%, greater than at least about 80%, greater than at least about 85%, or greater than at least about 90%, 95%, 96%, 97%, 98%, 99% or more (i.e. 100%) as compared to an identically sized fragment of a provided sequence. as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular). For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm. Global DNA sequence identity should be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an gap search with the following search parameters: gap open penalty, 12; and gap extension penalty, 1.

The subject nucleic acid compositions include full-length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of the polynucleotide sequences provided herein.

As discussed above, the polynucleotides useful in the methods described herein also include polynucleotide variants having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under high stringency conditions, for example, at 50°C or higher and 0.1XSSC (9 mM saline/0.9 mM sodium citrate). Hybridization methods and conditions are well known in the art, see, *e.g.*, USPN 5,707,829. Nucleic acids that are substantially identical to the provided polynucleotide sequences, *e.g.* allelic variants, genetically altered versions of the gene, *etc.*, bind to the provided polynucleotide sequences under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species, *e.g.* primate species, particularly human; rodents, such as rats and mice; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.*

In one embodiment, hybridization is performed using a fragment of at least 15 contiguous nucleotides (nt) of at least one of the polynucleotide sequences provided herein. That is, when at least 15 contiguous nt of one of the disclosed polynucleotide sequences is used as a probe, the probe will preferentially hybridize with a nucleic acid comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids that uniquely hybridize to the selected probe. Probes from more than one polynucleotide sequence provided herein can hybridize with the same nucleic acid if the cDNA from which they were derived corresponds to one mRNA.

Polynucleotides contemplated for use in the invention also include those having a sequence of naturally occurring variants of the nucleotide sequences (*e.g.*, degenerate variants (*e.g.*, sequences that encode the same polypeptides but, due to the degenerate nature of the genetic code, different in nucleotide sequence), allelic variants, *etc.*). Variants of the polynucleotides contemplated by the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the polynucleotides described herein can be identified where the allelic variant exhibits at most about 25-30% base pair (bp) mismatches relative to the selected polynucleotide probe. In general, allelic variants contain 15-25% bp mismatches, and can contain as little as even 5-15%, or 2-5%, or 1-2% bp mismatches, as well as a single bp mismatch.

The invention also encompasses homologs corresponding to any one of the polynucleotide sequences provided herein, where the source of homologous genes can be any mammalian species, *e.g.*, primate species, particularly human; rodents, such as rats; canines, felines, bovines, ovines, equines, yeast, nematodes, etc. Between mammalian species, *e.g.*, human and mouse, homologs generally have substantial sequence similarity, *e.g.*, at least 75% sequence identity, usually at least 80%%, at least 85, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about a fragment of a polynucleotide sequence and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known

in the art, such as gapped BLAST, described in Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402, or TeraBLAST available from TimeLogic Corp. (Crystal Bay, Nevada).

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein (e.g., in diagnosis, as a unique identifier of a differentially expressed gene of interest, etc.). The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide. mRNA species can also exist with both exons and introns, where the introns may be removed by alternative splicing. Furthermore it should be noted that different species of mRNAs encoded by the same genomic sequence can exist at varying levels in a cell, and detection of these various levels of mRNA species can be indicative of differential expression of the encoded gene product in the cell.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression.

The nucleic acid compositions of the subject invention can encode all or a part of the naturally-occurring polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.*

Probes specific to the polynucleotides described herein can be generated using the polynucleotide sequences disclosed herein. The probes are usually a fragment of a

polynucleotide sequences provided herein. The probes can be synthesized chemically or can be generated from longer polynucleotides using restriction enzymes. The probes can be labeled, for example, with a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of any one of the polynucleotide sequences provided herein. More preferably, probes are designed based on a contiguous sequence of one of the subject polynucleotides that remain unmasked following application of a masking program for masking low complexity (*e.g.*, XBLAST, RepeatMasker, etc.) to the sequence., *i.e.*, one would select an unmasked region, as indicated by the polynucleotides outside the poly-n stretches of the masked sequence produced by the masking program.

The polynucleotides of interest in the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the polynucleotides, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences that they are usually associated with , generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The polynucleotides described herein can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the polynucleotides can be regulated by their own or by other regulatory sequences known in the art. The polynucleotides can be introduced into suitable host cells using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

The nucleic acid compositions described herein can be used to, for example, produce polypeptides, as probes for the detection of mRNA in biological samples (*e.g.*, extracts of human cells) or cDNA produced from such samples, to generate additional copies of the polynucleotides, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example, determine the presence or absence of any one of the polynucleotide

provided herein or variants thereof in a sample. These and other uses are described in more detail below.

POLYPEPTIDES AND VARIANTS THEREOF

The present invention further provides polypeptides encoded by polynucleotides that represent genes that are differentially expressed in cancer cells. Such polypeptides are referred to herein as "polypeptides associated with cancer." The polypeptides can be used to generate antibodies specific for a polypeptide associated with cancer, which antibodies are in turn useful in diagnostic methods, prognostics methods, therametric methods, and the like as discussed in more detail herein. Polypeptides are also useful as targets for therapeutic intervention, as discussed in more detail herein.

The polypeptides contemplated by the invention include those encoded by the disclosed polynucleotides and the genes to which these polynucleotides correspond, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed polynucleotides. Further polypeptides contemplated by the invention include polypeptides that are encoded by polynucleotides that hybridize to polynucleotide of the sequence listing. Thus, the invention includes within its scope a polypeptide encoded by a polynucleotide having the sequence of any one of the polynucleotide sequences provided herein, or a variant thereof.

In general, the term "polypeptide" as used herein refers to both the full length polypeptide encoded by the recited polynucleotide, the polypeptide encoded by the gene represented by the recited polynucleotide, as well as portions or fragments thereof. "Polypeptides" also includes variants of the naturally occurring proteins, where such variants are homologous or substantially similar to the naturally occurring protein, and can be of an origin of the same or different species as the naturally occurring protein (*e.g.*, human, murine, or some other species that naturally expresses the recited polypeptide, usually a mammalian species). In general, variant polypeptides have a sequence that has at least about 80%, usually at least about 90%, and more usually at least about 98% sequence identity with a differentially expressed polypeptide described herein, as measured by BLAST 2.0 using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, *i.e.*, the polypeptide has

a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

The invention also encompasses homologs of the disclosed polypeptides (or fragments thereof) where the homologs are isolated from other species, *i.e.* other animal or plant species, where such homologs, usually mammalian species, *e.g.* rodents, such as mice, rats; domestic animals, *e.g.*, horse, cow, dog, cat; and humans. By "homolog" is meant a polypeptide having at least about 35%, usually at least about 40% and more usually at least about 60% amino acid sequence identity to a particular differentially expressed protein as identified above, where sequence identity is determined using the BLAST 2.0 algorithm, with the parameters described *supra.*

In general, the polypeptides of interest in the subject invention are provided in a non-naturally occurring environment, *e.g.* are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition that is enriched for the protein as compared to a cell or extract of a cell that naturally produces the protein. As such, isolated polypeptide is provided, where by "isolated" or "in substantially isolated form" is meant that the protein is present in a composition that is substantially free of other polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of other polypeptides of a cell that the protein is naturally found.

Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/ hydrophilicity, and/or steric bulk of the amino acid substituted.

Variants can be designed so as to retain or have enhanced biological activity of a particular region of the protein (*e.g.*, a functional domain and/or, where the polypeptide is a member of a protein family, a region associated with a consensus sequence). For example, muteins can be made which are optimized for increased antigenicity, i.e. amino acid variants of

a polypeptide may be made that increase the antigenicity of the polypeptide. Selection of amino acid alterations for production of variants can be based upon the accessibility (interior vs. exterior) of the amino acid (*see, e.g.,* Go *et al*, *Int. J. Peptide Protein Res.* (1980) *15*:211), the thermostability of the variant polypeptide (*see, e.g.,* Querol *et al.*, *Prot. Eng.* (1996) *9*:265), desired glycosylation sites (see, e.g., Olsen and Thomsen, *J. Gen. Microbiol.* (1991) *137*:579), desired disulfide bridges (*see, e.g.,* Clarke *et al.*, *Biochemistry* (1993) *32*:4322; and Wakarchuk *et al.*, *Protein Eng.* (1994) *7*:1379), desired metal binding sites (*see, e.g.,* Toma *et al.*, *Biochemistry* (1991) *30*:97, and Haezerbrouck *et al.*, *Protein Eng.* (1993) *6*:643), and desired substitutions with in proline loops (*see, e.g.,* Masul *et al.*, *Appl. Env. Microbiol.* (1994) *60*:3579). Cysteine-depleted muteins can be produced as disclosed in USPN 4,959,314.Variants also include fragments of the polypeptides disclosed herein, particularly biologically active fragments and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 aa to at least about 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a polynucleotide having a sequence of any one of the polynucleotide sequences provided herein, or a homolog thereof. The protein variants described herein are encoded by polynucleotides that are within the scope of the invention. The genetic code can be used to select the appropriate codons to construct the corresponding variants.

A fragment of a subject polypeptide is, for example, a polypeptide having an amino acid sequence which is a portion of a subject polypeptide e.g. a polypeptide encoded by a subject polynucleotide that is identified by any one of the sequence of SEQ ID NOS 1 - 499 or its complement. The polypeptide fragments of the invention are preferably at least about 9 aa, at least about 15 aa, and more preferably at least about 20 aa, still more preferably at least about 30 aa, and even more preferably, at least about 40 aa, at least about 50 aa, at least about 75 aa, at least about 100 aa, at least about 125 aa or at least about 150 aa in length. A fragment "at least 20 aa in length," for example, is intended to include 20 or more contiguous amino acids from, for example, the polypeptide encoded by a cDNA, in a cDNA clone contained in a deposited library, or a nucleotide sequence shown in SEQ ID NOS:1-23767 or the complementary stand thereof. In this context "about" includes the particularly recited value or a value larger or smaller by several (5, 4, 3, 2, or 1) amino acids. These

polypeptide fragments have uses that include, but are not limited to, production of antibodies as discussed herein. Of course, larger fragments (e.g., at least 150, 175, 200, 250, 500, 600, 1000, or 2000 amino acids in length) are also encompassed by the invention.

Moreover, representative examples of polypeptides fragments of the invention (useful in, for example, as antigens for antibody production), include, for example, fragments comprising, or alternatively consisting of, a sequence from about amino acid number 1-10, 5-10, 10-20, 21-31, 31-40, 41-61, 61-81, 91-120, 121-140, 141-162, 162-200, 201-240, 241-280, 281- 320, 321-360, 360-400, 400-450, 451-500, 500-600, 600-700, 700-800, 800-900 and the like. In this context "about" includes the particularly recited range or a range larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either terminus or at both termini. In some embodiments, these fragments has a functional activity (e.g., biological activity) whereas in other embodiments, these fragments may be used to make an antibody.

In one example, a polynucleotide having a sequence set forth in the sequence listing, containing no flanking sequences (i.e., consisting of the sequence set forth in the sequence listing), may be cloned into an expression vector having ATG and a stop codon (e.g. any one of the pET vector from Invitrogen, or other similar vectors from other manufactures), and used to express a polypeptide of interest encoded by the polynucleotide in a suitable cell, e.g., a bacterial cell. Accordingly, the polynucleotides may be used to produce polypeptides, and these polypeptides may be used to produce antibodies by known methods described above and below. In many embodiments, the sequence of the encoded polypeptide does not have to be known prior to its expression in a cell. However , if it desirable to know the sequence of the polypeptide, this may be derived from the sequence of the polynucleotide. Using the genetic code, the polynucleotide may be translated by hand, or by computer means. Suitable software for identifying open reading frames and translating them into polypeptide sequences are well know in the art, and include: Lasergene ™ from DNAStar (Madison, WI), and Vector NTI ™ from Informax (Frederick MD), and the like.

Further polypeptide variants may are described in PCT publications WO/00-55173, WO/01-07611 and WO/02-16429

VECTORS, HOST CELLS AND PROTEIN PRODUCTION

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides of the invention may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria.

Representative examples of appropriate hosts include,but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells (e.g., Saccharomyces cerevisiae or Pichia pastoris (ATCC Accession No. 201178)); insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. 5 Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNHSA; pNH16a,

pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRITS available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXTl and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Preferred expression vectors for use in yeast systems include, but are not limited to pYES2, pYDl, pTEFl/Zeo, pYES2/GS, pPICZ, pGAPZ, pGAPZalph, pPIC9, pPIC3.5, pHIL-D2, pHIL-Sl, pPIC3.5K, pPIC9K, and PAO815 (all available from Invitrogen, Carload, CA). Other suitable vectors will be readily apparent to the skilled artisan.

Nucleic acids of interest may be cloned into a suitable vector by route methods. Suitable vectors include plasmids, cosmids, recombinant viral vectors e.g. retroviral vectors, YACs, BACs and the like, phage vectors.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host mediated processes. Thus, it is well known

in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Suitable methods and compositions for polypeptide expression may be found in PCT publications WO/00-55173, WO/01-07611 and WO/02-16429, and suitable methods and compositions for production of modified polypeptides may be found in PCT publications WO/00-55173, WO/01-07611 and WO/02-16429.

ANTIBODIES AND OTHER POLYPEPTIDE OR POLYNUCLEOTIDE BINDING MOLECULES

The present invention further provides antibodies, which may be isolated antibodies, that are specific for a polypeptide encoded by a polynucleotide described herein and/or a polypeptide of a gene that corresponds to a polynucleotide described herein. Antibodies can be provided in a composition comprising the antibody and a buffer and/or a pharmaceutically acceptable excipient. Antibodies specific for a polypeptide associated with cancer are useful in a variety of diagnostic and therapeutic methods, as discussed in detail herein.

Gene products, including polypeptides, mRNA (particularly mRNAs having distinct secondary and/or tertiary structures), cDNA, or complete gene, can be prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. Antibodies may be used to identify a gene corresponding to a polynucleotide. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native protein in a cell or tissue preparation or in a cell-free extract of an in vitro expression system.

Antibodies

Further polypeptides of the invention relate to antibodies and T-cell antigen receptors (TCR) which immunospecifically bind a subject polypeptide, subject polypeptide fragment, or variant thereof, and/or an epitope thereof (as determined by immunoassays well known in the art for assaying specific antibody-antigen binding). Antibodies of the invention include, but are not limited to, polyclonal, monoclonal, multispecific, human, humanized or chimeric antibodies,

single chain antibodies, Fab fragments, F(ab') fragments, fragments produced by a Fab expression library, anti-idiotypic (anti-Id) antibodies (including, e.g., anti-Id antibodies to antibodies of the invention), and epitope-binding fragments of any of the above. The term "antibody," as used herein, refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site that immunospecifically binds an antigen. The immunoglobulin molecules of the invention can be of any type (e.g., IgG, IgE, IgM, IgD, IgA and IgY), class (e.g., IgG1, IgG2, lgG3, IgG4, IgAl and IgA2) or subclass of immunoglobulin molecule.

Most preferably the antibodies are human antigen-binding antibody fragments of the present invention and include, but are not limited to, Fab. Fab' and F(ab')2, Fd, single-chain Fvs (scFv), single-chain antibodies, disulfide-linked Fvs (sdFv) and fragments comprising either a V_L or V_H domain. Antigen-binding antibody fragments, including single-chain antibodies, may comprise the variable region(s) alone or in combination with the entirety or a portion of the following: hinge region, C_H1, C_H2, and C_H3 domains. Also included in the invention are antigen-binding fragments also comprising any combination of variable region(s) with a hinge region, C_H1, C_H2, and C_H3 domains. The antibodies of the invention may be from any animal origin including birds and mammals. Preferably, the antibodies are human, murine (e.g., mouse and rat), donkey, ship rabbit, goat, guinea pig, camel, horse, or chicken. As used herein, "human" antibodies include antibodies having the amino acid sequence of a human immunoglobulin and include antibodies isolated from, human immunoglobulin libraries or from animals transgenic for one or more human immunoglobulin and that do not express endogenous immunoglobulins, as described infra and, for example in, U.S. Patent No. 5,939,598 by Kucherlapati et al.

The antibodies of the present invention may be monospecific, bispecific, trispecific or of greater multispecificity. Multispecific antibodies may be specific for different epitopes of a polypeptide of the present invention or may be specific for both a polypeptide of the present invention as well as for a heterologous epitope, such as a heterologous polypeptide or solid support material. See, e.g., PCT publications WO 93/17715; WO 92/08802; WO 91/00360; WO 92/05793; Tutt, et al., J. Immunol. 147:60-69 (1991); U.S. Patent Nos. 4,474,893; 4,714,681; 4,925,648; 5,573,920; 5,601,819; Kostelny et al., J. Immunol. 148:1547-1553 (1992).

Antibodies of the present invention may be described or specified in terms of the epitope(s) or portion(s) of a polypeptide of the present invention which they recognize or

specifically bind. The epitope(s) or polypeptide portion(s) may be specified as described herein, e.g., by N-terminal and C-terminal positions, or by size in contiguous amino acid residues. Antibodies which specifically bind any epitope or polypeptide of the present invention may also be excluded. Therefore, the present invention includes antibodies that specifically bind polypeptides of the present invention, and allows for the exclusion of the same.

Antibodies of the present invention may also be described or specified in terms of their cross-reactivity. Antibodies that do not bind any other analog, ortholog, or homolog of a polypeptide of the present invention are included. Antibodies that bind polypeptides with at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 65%, at least 60%, at least 55%, and at least 50% identity (as calculated using methods known in the art and described herein) to a polypeptide of the present invention are also included in the present invention. In specific embodiments, antibodies of the present invention cross-react with murine, rat and/or rabbit homologs of human proteins and the corresponding epitopes thereof. Antibodies that do not bind polypeptides with less than 95%, less than 90%, less than 85%, less than 80%, less than 75%, less than 70%, less than 65%, less than 60%, less than 55%, and less than 50% identity (as calculated using methods known in the art and described herein) to a polypeptide of the present invention are also included in the present invention. In a specific embodiment, the above-described cross-reactivity is with respect to any single specific antigenic or immunogenic polypeptide, or combination(s) of 2, 3, 4, 5, or more of the specific antigenic and/or immunogenic polypeptides disclosed herein. Further included in the present invention are antibodies which bind polypeptides encoded by polynucleotides which hybridize to a polynucleotide of the present invention under stringent hybridization conditions (as described herein). Antibodies of the present invention may also be described or specified in terms of their binding affinity to a polypeptide of the invention. Preferred binding affinities include those with a dissociation constant or Kd less 5×10^{-5} M, 10^{-5} M, 5×10^{-6} M, 10^{-6}M, 5×10^{-7} M, 10^{-7} M, 5×10^{-8} M, 10^{-8} M, 5×10^{-9} M, 10^{-9} M, 5×10^{-10} M, 10-10 M, etc.

The invention also provides antibodies that competitively inhibit binding of an antibody to an epitope of the invention as determined by any method known in the art for determining competitive binding, for example, the immunoassays described herein. In preferred embodiments, the antibody competitively inhibits binding to the epitope by at least 95%, at least 90%, at least 85 %, at least 80%, at least 75%, at least 70%, at least 60%, or at least 50%.

Methods for making screening, assaying, humanizing, and modifying different types of antibody are well known in the art and may be found in PCT publications WO/00-55173, WO/01-07611 and WO/02-16429.

In addition, the invention further provides polynucleotides comprising a nucleotide sequence encoding an antibody of the invention and fragments thereof. The invention also encompasses polynucleotides that hybridize under stringent or alternatively, under lower stringency hybridization conditions, e.g., as defined supra, to polynucleotides that encode an antibody, preferably, that specifically binds to a polypeptide of the invention, preferably, an antibody that binds to a subject polypeptide.

The antibodies of the invention can be produced by any method known in the art for the synthesis of antibodies, in particular, by chemical synthesis or preferably, by recombinant expression techniques. Recombinant expression of an antibody of the invention, or fragment, derivative or analog thereof, (e.g., a heavy or light chain of an antibody of the invention or a single chain antibody of the invention), requires construction of an expression vector containing a polynucleotide that encodes the antibody. Once a polynucleotide encoding an antibody molecule or a heavy or light chain of an antibody, or portion thereof (preferably containing the heavy or light chain variable domain), of the invention has been obtained, the vector for the production of the antibody molecule may be produced by recombinant DNA technology using techniques well known in the art. Thus, methods for preparing a protein by expressing a polynucleotide containing an antibody encoding nucleotide sequence are described herein. Methods which are well known to those skilled in the art can be used to construct expression vectors containing antibody coding sequences and appropriate transcriptional and translational control signals. These methods include, for example, in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. The invention, thus, provides replicable vectors comprising a nucleotide sequence encoding an antibody molecule of the invention, or a heavy or light chain thereof, or a heavy or light chain variable domain, operably linked to a promoter. Such vectors may include the nucleotide sequence encoding the constant region of the antibody molecule (see, e.g., PCT Publication WO 86/05807; PCT Publication WO 89/01036; and U.S. Patent No. 5,122,464) and the variable domain of the antibody may be cloned into such a vector for expression of the entire heavy or light chain.

The expression vector is transferred to a host cell by conventional techniques and the transfected cells are then cultured by conventional techniques to produce an antibody of the invention. Thus, the invention includes host cells containing a polynucleotide encoding an antibody of the invention, or a heavy or light chain thereof, or a single chain antibody of the invention, operably linked to a heterologous promoter. In preferred embodiments for the expression of double-chained antibodies, vectors encoding both the heavy and light chains may be co-expressed in the host cell for expression of the entire immunoglobulin molecule, as detailed below.

A variety of host-expression vector systems may be utilized to express the antibody molecules of the invention. Such host-expression systems represent vehicles by which the coding sequences of interest may be produced and subsequently purified, but also represent cells which may, when transformed or transfected with the appropriate nucleotide coding sequences, express an antibody molecule of the invention in situ. These include but are not limited to microorganisms such as bacteria (e.g., E. coli, B. subtilis) transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing antibody coding sequences; yeast (e.g., Saccharomyces, Pichia) transformed with recombinant yeast expression vectors containing antibody coding sequences; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing antibody coding sequences; plant cell systems infected with recombinant virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing antibody coding sequences; or mammalian cell systems (e.g., COS, CHO, BHK, 293, 3T3 cells) harboring recombinant expression constructs containing promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter). Preferably, bacterial cells such as Escherichia coli, and more preferably, eukaryotic cells, especially for the expression of whole recombinant antibody molecule, are used for the expression of a recombinant antibody molecule. For example, mammalian cells such as Chinese hamster ovary cells (CHO), in conjunction with a vector such as the major intermediate early gene promoter element from human cytomegalovirus is an effective expression system for antibodies (Foecking et al., Gene 45:101 (1986); Cockett etal., Bio/Technology 8:2 (1990)).

Antibodies production is well known in the art. Exemplary methods and compositions for making antibodies may be found in PCT publications WO/00-55173, WO/01-07611 and WO/02-16429.

Immunophenotyping

The antibodies of the invention may be utilized for immunophenotyping of cell lines and biological samples. The translation product of the gene of the present invention may be useful as a cell specific marker, or more specifically as a cellular marker that is differentially expressed at various stages of differentiation and/or maturation of particular cell types. Monoclonal antibodies directed against a specific epitope, or combination of epitopes, will allow for the screening of cellular populations expressing the marker. Various techniques can be utilized using monoclonal antibodies to screen for cellular populations expressing the marker(s), and include magnetic separation using antibody-coated magnetic beads, "panning" with antibody attached to a solid matrix (i.e., plate), and flow cytometry (See, e.g., U.S. Patent 5,985,660; and Morrison et al. Cell, 96:737-49 (1999)).

These techniques allow for the screening of particular populations of cells, such as might be found with hematological malignancies (i.e. minimal residual disease (MRD) in acute leukemic patients) and "non-self cells in transplantations to prevent Graft-versus-Host Disease (GVHD). Alternatively, these techniques allow for the screening of hematopoietic stem and progenitor cells capable of undergoing proliferation and/or differentiation, as might be found in human umbilical cord blood.

KITS

Also provided by the subject invention are kits for practicing the subject methods, as described above. The subject kits include at least one or more of: a subject nucleic acid, isolated polypeptide or an antibody thereto. Other optional components of the kit include: restriction enzymes, control primers and plasmids; buffers, cells, carriers adjuvents etc. The nucleic acids of the kit may also have restrictions sites, multiple cloning sites, primer sites, etc to facilitate their ligation other plasmids. The various components of the kit may be present in separate containers or certain compatible components may be precombined into a single container, as desired. In many embodiments, kits with unit doses of the active agent, e.g. in oral or injectable doses, are provided. In certain embodiments, controls, such as samples from a cancerous or non-

cancerous cell are provided by the invention. Further embodiments of the kit include an antibody for a subject polypeptide and a chemotherapeutic agent to be used in combination with the polypeptide as a treatment.

In addition to above-mentioned components, the subject kits typically further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

COMPUTER-RELATED EMBODIMENTS

In general, a library of polynucleotides is a collection of sequence information, which information is provided in either biochemical form (*e.g.*, as a collection of polynucleotide molecules), or in electronic form (*e.g.*, as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The sequence information of the polynucleotides can be used in a variety of ways, *e.g.*, as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (*e.g.*, cell type markers), and/or as markers of a given disease or disease state. For example, in the instant case, the sequences of polynucleotides and polypeptides corresponding to genes differentially expressed in cancer, as well as the nucleic acid and amino acid sequences of the genes themselves, can be provided in electronic form in a computer database.

In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (*e.g.*, a cell of the same or similar type that is not substantially affected by disease). For example, a

polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either overexpressed or underexpressed in a cancerous cell affected by cancer relative to a normal (*i.e.*, substantially disease-free) cell.

The nucleotide sequence information of the library can be embodied in any suitable form, *e.g.*, electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file (or, in biochemical form, a collection of nucleic acid molecules) that contains the representative nucleotide sequences of genes that are differentially expressed (*e.g.*, overexpressed or underexpressed) as between, for example, i) a cancerous cell and a normal cell; ii) a cancerous cell and a dysplastic cell; iii) a cancerous cell and a cell affected by a disease or condition other than cancer; iv) a metastatic cancerous cell and a normal cell and/or non-metastatic cancerous cell; v) a malignant cancerous cell and a non-malignant cancerous cell (or a normal cell) and/or vi) a dysplastic cell relative to a normal cell. Other combinations and comparisons of cells affected by various diseases or stages of disease will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acids that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

The polynucleotide libraries of the subject invention generally comprise sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence of any of sequence described herein. By plurality is meant at least 2, usually at least 3 and can include up to all of the sequences described herein. The length and number of polynucleotides in the library will vary with the nature of the library, *e.g.*, if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, etc.

Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. “Media” refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, *e.g.* the nucleic acid sequences of any of the polynucleotides of the sequences described herein, can be recorded on computer readable

media, *e.g.* any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, *e.g.* word processing text file, database format, *etc.* In addition to the sequence information, electronic versions of libraries comprising one or more sequence described herein can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (*e.g.*, searchable files, executable files, *etc*, including, but not limited to, for example, search program software, *etc.*).

By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information (e.g. the NCBI sequence database) is publicly available. For example, the gapped BLAST (Altschul *et al., Nucleic Acids Res.* (1997) *25*:3389-3402) and BLAZE (Brutlag *et al., Comp. Chem.* (1993) 17:203) search algorithms on a Sybase system, or the TeraBLAST (TimeLogic, Crystal Bay, Nevada) program optionally running on a specialized computer platform available from TimeLogic, can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means

can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

"Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, *e.g.* MacPattern (EMBL), TeraBLAST (TimeLogic), BLASTN and BLASTX (NCBI). A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nt. A variety of means for comparing nucleic acids or polypeptides may be used to compare accomplish a sequence comparison (e.g., to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used to search the computer based systems of the present invention to compare of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art.

A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences, kinase domains, receptor binding domains, SH2 domains, SH3 domains, phosphorylation sites, protein interaction domains, transmembrane domains, etc. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile. A gene expression profile can be generated from, for example, a cDNA

library prepared from mRNA isolated from a test cell suspected of being cancerous or pre-cancerous, comparing the sequences or partial sequences of the clones against the sequences in an electronic database, where the sequences of the electronic database represent genes differentially expressed in a cancerous cell, *e.g.,* a cancerous breast cell. The number of clones having a sequence that has substantial similarity to a sequence that represents a gene differentially expressed in a cancerous cell is then determined, and the number of clones corresponding to each of such genes is determined. An increased number of clones that correspond to differentially expressed gene is present in the cDNA library of the test cell (relative to, for example, the number of clones expected in a cDNA of a normal cell) indicates that the test cell is cancerous.

As discussed above, the "library" as used herein also encompasses biochemical libraries of the polynucleotides of the sequences described herein, *e.g.*, collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, *e.g.*, a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (*i.e.*, an array) and the like. Of particular interest are nucleic acid arrays in which one or more of the genes described herein is represented by a sequence on the array. By array is meant an article of manufacture that has at least a substrate with at least two distinct nucleic acid targets on one of its surfaces, where the number of distinct nucleic acids can be considerably higher, typically being at least 10 nt, usually at least 20 nt and often at least 25 nt. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the polypeptides of the library will represent at least a portion of the polypeptides encoded by a gene corresponding to a sequence described herein.

DIAGNOSTIC AND OTHER METHODS INVOLVING DETECTION OF DIFFERENTIALLY EXPRESSED GENES

The present invention provides methods of using the polynucleotides described herein in, for example, diagnosis of cancer and classification of cancer cells according to expression

profiles. In specific non-limiting embodiments, the methods are useful for detecting cancer cells, facilitating diagnosis of cancer and the severity of a cancer (*e.g.*, tumor grade, tumor burden, and the like) in a subject, facilitating a determination of the prognosis of a subject, and assessing the responsiveness of the subject to therapy (*e.g.*, by providing a measure of therapeutic effect through, for example, assessing tumor burden during or following a chemotherapeutic regimen). Detection can be based on detection of a polynucleotide that is differentially expressed in a cancer cell, and/or detection of a polypeptide encoded by a polynucleotide that is differentially expressed in a cancer cell ("a polypeptide associated with cancer"). The detection methods of the invention can be conducted *in vitro* or *in vivo*, on isolated cells, or in whole tissues or a bodily fluid, *e.g.*, blood, plasma, serum, urine, and the like).

In general, methods of the invention involving detection of a gene product (*e.g.*, mRNA, cDNA generated from such mRNA, and polypeptides) involve contacting a sample with a probe specific for the gene product of interest. "Probe" as used herein in such methods is meant to refer to a molecule that specifically binds a gene product of interest (*e.g.*, the probe binds to the target gene product with a specificity sufficient to distinguish binding to target over non-specific binding to non-target (background) molecules). "Probes" include, but are not necessarily limited to, nucleic acid probes (*e.g.*, DNA, RNA, modified nucleic acid, and the like), antibodies (*e.g.*, antibodies, antibody fragments that retain binding to a target epitope, single chain antibodies, and the like), or other polypeptide, peptide, or molecule (*e.g.*, receptor ligand) that specifically binds a target gene product of interest.

The probe and sample suspected of having the gene product of interest are contacted under conditions suitable for binding of the probe to the gene product. For example, contacting is generally for a time sufficient to allow binding of the probe to the gene product (*e.g.*, from several minutes to a few hours), and at a temperature and conditions of osmolarity and the like that provide for binding of the probe to the gene product at a level that is sufficiently distinguishable from background binding of the probe (*e.g.*, under conditions that minimize non-specific binding). Suitable conditions for probe-target gene product binding can be readily determined using controls and other techniques available and known to one of ordinary skill in the art.

In this embodiment, the probe can be an antibody or other polypeptide, peptide, or molecule (*e.g.*, receptor ligand) that specifically binds a target polypeptide of interest.

The detection methods can be provided as part of a kit. Thus, the invention further provides kits for detecting the presence and/or a level of a polynucleotide that is differentially expressed in a cancer cell (*e.g.*, by detection of an mRNA encoded by the differentially expressed gene of interest), and/or a polypeptide encoded thereby, in a biological sample. Procedures using these kits can be performed by clinical laboratories, experimental laboratories, medical practitioners, or private individuals. The kits of the invention for detecting a polypeptide encoded by a polynucleotide that is differentially expressed in a cancer cell comprise a moiety that specifically binds the polypeptide, which may be a specific antibody. The kits of the invention for detecting a polynucleotide that is differentially expressed in a cancer cell comprise a moiety that specifically hybridizes to such a polynucleotide. The kit may optionally provide additional components that are useful in the procedure, including, but not limited to, buffers, developing reagents, labels, reacting surfaces, means for detection, control samples, standards, instructions, and interpretive information.

<u>Detecting a polypeptide encoded by a polynucleotide that is differentially expressed in a cancer cell</u>

In some embodiments, methods are provided for a detecting cancer cell by detecting in a cell, a polypeptide encoded by a gene differentially expressed in a cancer cell. Any of a variety of known methods can be used for detection, including, but not limited to, immunoassay, using an antibody specific for the encoded polypeptide, e.g., by enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and the like; and functional assays for the encoded polypeptide, e.g., binding activity or enzymatic activity.

For example, an immunofluorescence assay can be easily performed on cells without first isolating the encoded polypeptide. The cells are first fixed onto a solid support, such as a microscope slide or microtiter well. This fixing step can permeabilize the cell membrane. The permeablization of the cell membrane permits the polypeptide-specific probe (*e.g*, antibody) to bind. Alternatively, where the polypeptide is secreted or membrane-bound, or is otherwise accessible at the cell-surface (*e.g.*, receptors, and other molecule stably-associated with the outer cell membrane or otherwise stably associated with the cell membrane, such permeabilization may not be necessary.

Next, the fixed cells are exposed to an antibody specific for the encoded polypeptide. To increase the sensitivity of the assay, the fixed cells may be further exposed to a second antibody, which is labeled and binds to the first antibody, which is specific for the encoded polypeptide. Typically, the secondary antibody is detectably labeled, e.g., with a fluorescent marker. The cells which express the encoded polypeptide will be fluorescently labeled and easily visualized under the microscope. See, for example, Hashido *et al.* (1992) *Biochem. Biophys. Res. Comm.* 187:1241-1248.

As will be readily apparent to the ordinarily skilled artisan upon reading the present specification, the detection methods and other methods described herein can be varied. Such variations are within the intended scope of the invention. For example, in the above detection scheme, the probe for use in detection can be immobilized on a solid support, and the test sample contacted with the immobilized probe. Binding of the test sample to the probe can then be detected in a variety of ways, *e.g.*, by detecting a detectable label bound to the test sample.

The present invention further provides methods for detecting the presence of and/or measuring a level of a polypeptide in a biological sample, which polypeptide is encoded by a polynucleotide that represents a gene differentially expressed in cancer, particularly in a polynucleotide that represents a gene differentially cancer cell, using a probe specific for the encoded polypeptide. In this embodiment, the probe can be a an antibody or other polypeptide, peptide, or molecule (*e.g.*, receptor ligand) that specifically binds a target polypeptide of interest.

The methods generally comprise: a) contacting the sample with an antibody specific for a differentially expressed polypeptide in a test cell; and b) detecting binding between the antibody and molecules of the sample. The level of antibody binding (either qualitative or quantitative) indicates the cancerous state of the cell. For example, where the differentially expressed gene is increased in cancerous cells, detection of an increased level of antibody binding to the test sample relative to antibody binding level associated with a normal cell indicates that the test cell is cancerous.

Suitable controls include a sample known not to contain the encoded polypeptide; and a sample contacted with an antibody not specific for the encoded polypeptide, e.g., an anti-idiotype antibody. A variety of methods to detect specific antibody-antigen interactions are known in the art and can be used in the method, including, but not limited to, standard

immunohistological methods, immunoprecipitation, an enzyme immunoassay, and a radioimmunoassay.

In general, the specific antibody will be detectably labeled, either directly or indirectly. Direct labels include radioisotopes; enzymes whose products are detectable (e.g., luciferase, β-galactosidase, and the like); fluorescent labels (e.g., fluorescein isothiocyanate, rhodamine, phycoerythrin, and the like); fluorescence emitting metals, e.g., ^{152}Eu, or others of the lanthanide series, attached to the antibody through metal chelating groups such as EDTA; chemiluminescent compounds, e.g., luminol, isoluminol, acridinium salts, and the like; bioluminescent compounds, e.g., luciferin, aequorin (green fluorescent protein), and the like.

The antibody may be attached (coupled) to an insoluble support, such as a polystyrene plate or a bead. Indirect labels include second antibodies specific for antibodies specific for the encoded polypeptide ("first specific antibody"), wherein the second antibody is labeled as described above; and members of specific binding pairs, e.g., biotin-avidin, and the like. The biological sample may be brought into contact with and immobilized on a solid support or carrier, such as nitrocellulose, that is capable of immobilizing cells, cell particles, or soluble proteins. The support may then be washed with suitable buffers, followed by contacting with a detectably-labeled first specific antibody. Detection methods are known in the art and will be chosen as appropriate to the signal emitted by the detectable label. Detection is generally accomplished in comparison to suitable controls, and to appropriate standards.

In some embodiments, the methods are adapted for use *in vivo*, e.g., to locate or identify sites where cancer cells are present. In these embodiments, a detectably-labeled moiety, e.g., an antibody, which is specific for a cancer-associated polypeptide is administered to an individual (e.g., by injection), and labeled cells are located using standard imaging techniques, including, but not limited to, magnetic resonance imaging, computed tomography scanning, and the like. In this manner, cancer cells are differentially labeled.

Detecting a polynucleotide that represents a gene differentially expressed in a cancer cell

In some embodiments, methods are provided for detecting a cancer cell by detecting expression in the cell of a transcript or that is differentially expressed in a cancer cell. Any of a variety of known methods can be used for detection, including, but not limited to, detection of a transcript by hybridization with a polynucleotide that hybridizes to a polynucleotide that is differentially expressed in a cancer cell; detection of a transcript by a polymerase chain reaction

using specific oligonucleotide primers; *in situ* hybridization of a cell using as a probe a polynucleotide that hybridizes to a gene that is differentially expressed in a cancer cell and the like.

In many embodiments, the levels of a subject gene product are measured. By measured is meant qualitatively or quantitatively estimating the level of the gene product in a first biological sample either directly (e.g. by determining or estimating absolute levels of gene product) or relatively by comparing the levels to a second control biological sample. In many embodiments the second control biological sample is obtained from an individual not having not having cancer. As will be appreciated in the art, once a standard control level of gene expression is known, it can be used repeatedly as a standard for comparison. Other control samples include samples of cancerous tissue.

The methods can be used to detect and/or measure mRNA levels of a gene that is differentially expressed in a cancer cell. In some embodiments, the methods comprise: a) contacting a sample with a polynucleotide that corresponds to a differentially expressed gene described herein under conditions that allow hybridization; and b) detecting hybridization, if any. Detection of differential hybridization, when compared to a suitable control, is an indication of the presence in the sample of a polynucleotide that is differentially expressed in a cancer cell. Appropriate controls include, for example, a sample that is known not to contain a polynucleotide that is differentially expressed in a cancer cell. Conditions that allow hybridization are known in the art, and have been described in more detail above.

Detection can also be accomplished by any known method, including, but not limited to, *in situ* hybridization, PCR (polymerase chain reaction), RT-PCR (reverse transcription-PCR), and "Northern" or RNA blotting, arrays, microarrays, etc, or combinations of such techniques, using a suitably labeled polynucleotide. A variety of labels and labeling methods for polynucleotides are known in the art and can be used in the assay methods of the invention. Specific hybridization can be determined by comparison to appropriate controls.

Polynucleotides described herein are used for a variety of purposes, such as probes for detection of and/or measurement of, transcription levels of a polynucleotide that is differentially expressed in a cancer cell. Additional disclosure about preferred regions of the disclosed polynucleotide sequences is found in the Examples. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 2-, 5-, 10-, or 20-fold

higher than the background hybridization provided with other unrelated sequences. It should be noted that "probe" as used in this context of detection of nucleic acid is meant to refer to a polynucleotide sequence used to detect a differentially expressed gene product in a test sample. As will be readily appreciated by the ordinarily skilled artisan, the probe can be detectably labeled and contacted with, for example, an array comprising immobilized polynucleotides obtained from a test sample (*e.g.*, mRNA). Alternatively, the probe can be immobilized on an array and the test sample detectably labeled. These and other variations of the methods of the invention are well within the skill in the art and are within the scope of the invention.

Labeled nucleic acid probes may be used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots, mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization can be quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for in situ hybridization to cells to detect expression. Probes can also be used *in vivo* for diagnostic detection of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores, fluorophores, and enzymes. Other examples of nucleotide hybridization assays are described in WO92/02526 and USPN 5,124,246.

PCR is another means for detecting small amounts of target nucleic acids, methods for which may be found in Sambrook, *et al.* Molecular Cloning: A Laboratory Manual, CSH Press 1989, pp.14.2-14.33.

A detectable label may be included in the amplification reaction. Suitable detectable labels include fluorochromes,(*e.g.* fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)), radioactive labels, (*e.g.* ^{32}P, ^{35}S, ^{3}H, *etc.*), and the like. The label may be a two stage system, where the polynucleotides is conjugated to biotin, haptens, *etc.* having a high affinity binding partner, *e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the

primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

Arrays

Polynucleotide arrays provide a high throughput technique that can assay a large number of polynucleotides or polypeptides in a sample. This technology can be used as a tool to test for differential expression.

A variety of methods of producing arrays, as well as variations of these methods, are known in the art and contemplated for use in the invention. For example, arrays can be created by spotting polynucleotide probes onto a substrate (*e.g.*, glass, nitrocellulose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions.

Samples of polynucleotides can be detectably labeled (*e.g.*, using radioactive or fluorescent labels) and then hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Alternatively, the polynucleotides of the test sample can be immobilized on the array, and the probes detectably labeled. Techniques for constructing arrays and methods of using these arrays are described in, for example, Schena *et al.* (1996) *Proc Natl Acad Sci* U S A. 93(20):10614-9; Schena *et al.* (1995) *Science* 270(5235):467-70; Shalon *et al.* (1996) *Genome Res.* 6(7):639-45, USPN 5,807,522, EP 799 897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; USPN 5,593,839; USPN 5,578,832; EP 728 520; USPN 5,599,695; EP 721 016; USPN 5,556,752; WO 95/22058; and USPN 5,631,734. In most embodiments, the "probe" is detectably labeled. In other embodiments, the probe is immobilized on the array and not detectably labeled.

Arrays can be used, for example, to examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential expression of a gene corresponding to a polynucleotide described herein, where expression is compared between a test cell and control cell (*e.g.,* cancer cells and normal cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific gene product. Exemplary uses of arrays are further described in, for example, Pappalarado *et al.*, *Sem. Radiation Oncol.* (1998) *8*:217; and Ramsay, *Nature Biotechnol.* (1998) *16*:40. Furthermore, many variations on

methods of detection using arrays are well within the skill in the art and within the scope of the present invention. For example, rather than immobilizing the probe to a solid support, the test sample can be immobilized on a solid support which is then contacted with the probe.

DIAGNOSIS, PROGNOSIS, ASSESSMENT OF THERAPY (THERAMETRICS), AND MANAGEMENT OF CANCER

The polynucleotides described herein, as well as their gene products and corresponding genes and gene products, are of particular interest as genetic or biochemical markers (e.g., in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or to monitor the efficacy of various therapies and preventive interventions.

For example, the level of expression of certain polynucleotides can be indicative of a poorer prognosis, and therefore warrant more aggressive chemo- or radio-therapy for a patient or vice versa. The correlation of novel surrogate tumor specific features with response to treatment and outcome in patients can define prognostic indicators that allow the design of tailored therapy based on the molecular profile of the tumor. These therapies include antibody targeting, antagonists (*e.g.*, small molecules), and gene therapy.

Determining expression of certain polynucleotides and comparison of a patient's profile with known expression in normal tissue and variants of the disease allows a determination of the best possible treatment for a patient, both in terms of specificity of treatment and in terms of comfort level of the patient. Surrogate tumor markers, such as polynucleotide expression, can also be used to better classify, and thus diagnose and treat, different forms and disease states of cancer. Two classifications widely used in oncology that can benefit from identification of the expression levels of the genes corresponding to the polynucleotides described herein are staging of the cancerous disorder, and grading the nature of the cancerous tissue.

The polynucleotides that correspond to differentially expressed genes, as well as their encoded gene products, can be useful to monitor patients having or susceptible to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. In addition, the polynucleotides described herein, as well as the genes corresponding to such polynucleotides, can be useful as therametrics, *e.g.*, to assess the effectiveness of therapy by using the polynucleotides or their encoded gene products, to assess, for example, tumor burden in the patient before, during, and after therapy.

Furthermore, a polynucleotide identified as corresponding to a gene that is differentially expressed in, and thus is important for, one type of cancer can also have implications for development or risk of development of other types of cancer, e.g., where a polynucleotide represents a gene differentially expressed across various cancer types. Thus, for example, expression of a polynucleotide corresponding to a gene that has clinical implications for cancer can also have clinical implications for metastatic breast cancer, colon cancer, or ovarian cancer, etc.

Staging. Staging is a process used by physicians to describe how advanced the cancerous state is in a patient. Staging assists the physician in determining a prognosis, planning treatment and evaluating the results of such treatment. Staging systems vary with the types of cancer, but generally involve the following "TNM" system: the type of tumor, indicated by T; whether the cancer has metastasized to nearby lymph nodes, indicated by N; and whether the cancer has metastasized to more distant parts of the body, indicated by M. Generally, if a cancer is only detectable in the area of the primary lesion without having spread to any lymph nodes it is called Stage I. If it has spread only to the closest lymph nodes, it is called Stage II. In Stage III, the cancer has generally spread to the lymph nodes in near proximity to the site of the primary lesion. Cancers that have spread to a distant part of the body, such as the liver, bone, brain or other site, are Stage IV, the most advanced stage.

The polynucleotides and corresponding genes and gene products described herein can facilitate fine-tuning of the staging process by identifying markers for the aggressiveness of a cancer, *e.g.* the metastatic potential, as well as the presence in different areas of the body. Thus, a Stage II cancer with a polynucleotide signifying a high metastatic potential cancer can be used to change a borderline Stage II tumor to a Stage III tumor, justifying more aggressive therapy. Conversely, the presence of a polynucleotide signifying a lower metastatic potential allows more conservative staging of a tumor.

One type of breast cancer is ductal carcinoma in situ (DCIS): DCIS is when the breast cancer cells are completely contained within the breast ducts (the channels in the breast that carry milk to the nipple), and have not spread into the surrounding breast tissue. This may also be referred to as non-invasive or intraductal cancer, as the cancer cells have not yet spread into the surrounding breast tissue and so usually have not spread into any other part of the body.

Lobular carcinoma in situ breast cancer (LCIS) means that cell changes are found in the lining of the lobules of the breast. It can be present in both breasts. It is also referred to as non-invasive cancer as it has not spread into the surrounding breast tissue.

Invasive breast cancer can be staged as follows: Stage 1 tumours: these measure less than two centimetres. The lymph glands in the armpit are not affected and there are no signs that the cancer has spread elsewhere in the body; Stage 2 tumours: these measure between two and five centimetres, or the lymph glands in the armpit are affected, or both. However, there are no signs that the cancer has spread further; Stage 3 tumours: these are larger than five centimetres and may be attached to surrounding structures such as the muscle or skin. The lymph glands are usually affected, but there are no signs that the cancer has spread beyond the breast or the lymph glands in the armpit; Stage 4 tumours: these are of any size, but the lymph glands are usually affected and the cancer has spread to other parts of the body. This is secondary breast cancer.

<u>Grading of cancers.</u> Grade is a term used to describe how closely a tumor resembles normal tissue of its same type. The microscopic appearance of a tumor is used to identify tumor grade based on parameters such as cell morphology, cellular organization, and other markers of differentiation. As a general rule, the grade of a tumor corresponds to its rate of growth or aggressiveness, with undifferentiated or high-grade tumors generally being more aggressive than well-differentiated or low-grade tumors.

The polynucleotides of the Sequence Listing, and their corresponding genes and gene products, can be especially valuable in determining the grade of the tumor, as they not only can aid in determining the differentiation status of the cells of a tumor, they can also identify factors other than differentiation that are valuable in determining the aggressiveness of a tumor, such as metastatic potential.

Low grade means that the cancer cells look very like the normal cells. They are usually slowly growing and are less likely to spread. In high grade tumors the cells look very abnormal. They are likely to grow more quickly and are more likely to spread.

<u>Assessment of proliferation of cells in tumor.</u> The differential expression level of the polynucleotides described herein can facilitate assessment of the rate of proliferation of tumor cells, and thus provide an indicator of the aggressiveness of the rate of tumor growth. For example, assessment of the relative expression levels of genes involved in cell cycle can provide an indication of cellular proliferation, and thus serve as a marker of proliferation.

Detection of cancer.

The polynucleotides corresponding to genes that exhibit the appropriate expression pattern can be used to detect cancer in a subject. The expression of appropriate polynucleotides can be used in the diagnosis, prognosis and management of cancer. Detection of cancer can be determined using expression levels of any of these sequences alone or in combination with the levels of expression of other known cancer genes. Determination of the aggressive nature and/or the metastatic potential of a cancer can be determined by comparing levels of one or more gene products of the genes corresponding to the polynucleotides described herein, and comparing total levels of another sequence known to vary in cancerous tissue, *e.g.*, expression of p53, DCC, ras, FAP (see, e.g., Fearon ER, *et al.*, *Cell* (1990) *61(5)*:759; Hamilton SR *et al.*, *Cancer* (1993) *72*:957; Bodmer W, *et al.*, *Nat Genet.* (1994) *4(3)*:217; Fearon ER, *Ann N Y Acad Sci.* (1995) *768*:101). For example, development of cancer can be detected by examining the level of expression of a gene corresponding to a polynucleotides described herein to the levels of oncogenes (*e.g.* ras) or tumor suppressor genes (*e.g.* FAP or p53). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous tissue, to discriminate between cancers with different cells of origin, to discriminate between cancers with different potential metastatic rates, etc. For a review of other markers of cancer, see, *e.g.*, Hanahan et al. (2000) Cell 100:57-70.

Treatment of cancer

The invention further provides methods for reducing growth of cancer cells. The methods provide for decreasing the expression of a gene that is differentially expressed in a cancer cell or decreasing the level of and/or decreasing an activity of a cancer-associated polypeptide. In general, the methods comprise contacting a cancer cell with a substance that modulates (1) expression of a gene that is differentially expressed in cancer; or (2) a level of and/or an activity of a cancer-associated polypeptide.

"Reducing growth of cancer cells" includes, but is not limited to, reducing proliferation of cancer cells, and reducing the incidence of a non-cancerous cell becoming a cancerous cell. Whether a reduction in cancer cell growth has been achieved can be readily determined using any known assay, including, but not limited to, [^{3}H]-thymidine incorporation; counting cell

number over a period of time; detecting and/or measuring a marker associated with breast cancer (e.g., PSA).

The present invention provides methods for treating cancer, generally comprising administering to an individual in need thereof a substance that reduces cancer cell growth, in an amount sufficient to reduce cancer cell growth and treat the cancer. Whether a substance, or a specific amount of the substance, is effective in treating cancer can be assessed using any of a variety of known diagnostic assays for cancer, including, but not limited to, proctoscopy, rectal examination, biopsy, contrast radiographic studies, CAT scan, and detection of a tumor marker associated with cancer in the blood of the individual (*e.g.*, PSA (breast-specific antigen)). The substance can be administered systemically or locally. Thus, in some embodiments, the substance is administered locally, and cancer growth is decreased at the site of administration. Local administration may be useful in treating, e.g., a solid tumor.

A substance that reduces cancer cell growth can be targeted to a cancer cell. Thus, in some embodiments, the invention provides a method of delivering a drug to a cancer cell, comprising administering a drug-antibody complex to a subject, wherein the antibody is specific for a cancer-associated polypeptide, and the drug is one that reduces cancer cell growth, a variety of which are known in the art. Targeting can be accomplished by coupling (e.g., linking, directly or via a linker molecule, either covalently or non-covalently, so as to form a drug-antibody complex) a drug to an antibody specific for a cancer-associated polypeptide. Methods of coupling a drug to an antibody are well known in the art and need not be elaborated upon herein.

Tumor classification and patient stratification

The invention further provides for methods of classifying tumors, and thus grouping or "stratifying" patients, according to the expression profile of selected differentially expressed genes in a tumor. Differentially expressed genes can be analyzed for correlation with other differentially expressed genes in a single tumor type or across tumor types. Genes that demonstrate consistent correlation in expression profile in a given cancer cell type (*e.g.*, in a cancer cell or type of cancer) can be grouped together, e.g., when one gene is overexpressed in a tumor, a second gene is also usually overexpressed. Tumors can then be classified according to the expression profile of one or more genes selected from one or more groups.

The tumor of each patient in a pool of potential patients can be classified as described above. Patients having similarly classified tumors can then be selected for participation in an investigative or clinical trial of a cancer therapeutic where a homogeneous population is desired. The tumor classification of a patient can also be used in assessing the efficacy of a cancer therapeutic in a heterogeneous patient population. In addition, therapy for a patient having a tumor of a given expression profile can then be selected accordingly.

In another embodiment, differentially expressed gene products (*e.g.*, polypeptides or polynucleotides encoding such polypeptides) may be effectively used in treatment through vaccination. The growth of cancer cells is naturally limited in part due to immune surveillance. Stimulation of the immune system using a particular tumor-specific antigen enhances the effect towards the tumor expressing the antigen. An active vaccine comprising a polypeptide encoded by the cDNA of this invention would be appropriately administered to subjects having an alteration, e.g., overabundance, of the corresponding RNA, or those predisposed for developing cancer cells with an alteration of the same RNA. Polypeptide antigens are typically combined with an adjuvant as part of a vaccine composition. The vaccine is preferably administered first as a priming dose, and then again as a boosting dose, usually at least four weeks later. Further boosting doses may be given to enhance the effect. The dose and its timing are usually determined by the person responsible for the treatment.

The invention also encompasses the selection of a therapeutic regimen based upon the expression profile of differentially expressed genes in the patient's tumor. For example, a tumor can be analyzed for its expression profile of the genes corresponding to SEQ ID NOS:1-23767 as described herein, *e.g.*, the tumor is analyzed to determine which genes are expressed at elevated levels or at decreased levels relative to normal cells of the same tissue type. The expression patterns of the tumor are then compared to the expression patterns of tumors that respond to a selected therapy. Where the expression profiles of the test tumor cell and the expression profile of a tumor cell of known drug responsivity at least substantially match (*e.g.*, selected sets of genes at elevated levels in the tumor of known drug responsivity and are also at elevated levels in the test tumor cell), then the therapeutic agent selected for therapy is the drug to which tumors with that expression pattern respond.

<u>Pattern matching in diagnosis using arrays</u>

In another embodiment, the diagnostic and/or prognostic methods of the invention involve detection of expression of a selected set of genes in a test sample to produce a test expression pattern (TEP). The TEP is compared to a reference expression pattern (REP), which is generated by detection of expression of the selected set of genes in a reference sample (*e.g.*, a positive or negative control sample). The selected set of genes includes at least one of the genes of the invention, which genes correspond to the polynucleotide sequences described herein. Of particular interest is a selected set of genes that includes gene differentially expressed in the disease for which the test sample is to be screened.

IDENTIFICATION OF THERAPEUTIC TARGETS AND ANTI-CANCER THERAPEUTIC AGENTS

The present invention also encompasses methods for identification of agents having the ability to modulate activity of a differentially expressed gene product, as well as methods for identifying a differentially expressed gene product as a therapeutic target for treatment of cancer.

Identification of compounds that modulate activity of a differentially expressed gene product can be accomplished using any of a variety of drug screening techniques. Such agents are candidates for development of cancer therapies. Of particular interest are screening assays for agents that have tolerable toxicity for normal, non-cancerous human cells. The screening assays of the invention are generally based upon the ability of the agent to modulate an activity of a differentially expressed gene product and/or to inhibit or suppress phenomenon associated with cancer (*e.g.*, cell proliferation, colony formation, cell cycle arrest, metastasis, and the like).

Screening of candidate agents

Screening assays can be based upon any of a variety of techniques readily available and known to one of ordinary skill in the art. In general, the screening assays involve contacting a cancerous cell with a candidate agent, and assessing the effect upon biological activity of a differentially expressed gene product. The effect upon a biological activity can be detected by, for example, detection of expression of a gene product of a differentially expressed gene (*e.g.*, a decrease in mRNA or polypeptide levels, would in turn cause a decrease in biological activity of the gene product). Alternatively or in addition, the effect of the candidate agent can be assessed by examining the effect of the candidate agent in a functional assay. For example, where the differentially expressed gene product is an enzyme, then the effect upon biological activity can

be assessed by detecting a level of enzymatic activity associated with the differentially expressed gene product. The functional assay will be selected according to the differentially expressed gene product. In general, where the differentially expressed gene is increased in expression in a cancerous cell, agents of interest are those that decrease activity of the differentially expressed gene product.

Assays described infra can be readily adapted in the screening assay embodiments of the invention. Exemplary assays useful in screening candidate agents include, but are not limited to, hybridization-based assays (*e.g.*, use of nucleic acid probes or primers to assess expression levels), antibody-based assays (*e.g.*, to assess levels of polypeptide gene products), binding assays (*e.g.*, to detect interaction of a candidate agent with a differentially expressed polypeptide, which assays may be competitive assays where a natural or synthetic ligand for the polypeptide is available), and the like. Additional exemplary assays include, but are not necessarily limited to, cell proliferation assays, antisense knockout assays, assays to detect inhibition of cell cycle, assays of induction of cell death/apoptosis, and the like. Generally such assays are conducted *in vitro*, but many assays can be adapted for *in vivo* analyses, *e.g.*, in an animal model of the cancer.

Identification of therapeutic targets

In another embodiment, the invention contemplates identification of differentially expressed genes and gene products as therapeutic targets. In some respects, this is the converse of the assays described above for identification of agents having activity in modulating (*e.g.*, decreasing or increasing) activity of a differentially expressed gene product.

In this embodiment, therapeutic targets are identified by examining the effect(s) of an agent that can be demonstrated or has been demonstrated to modulate a cancerous phenotype (*e.g.*, inhibit or suppress or prevent development of a cancerous phenotype). Such agents are generally referred to herein as an "anti-cancer agent", which agents encompass chemotherapeutic agents. For example, the agent can be an antisense oligonucleotide that is specific for a selected gene transcript. For example, the antisense oligonucleotide may have a sequence corresponding to a sequence of a differentially expressed gene described herein, *e.g.*, a sequence of one of SEQ ID NOS:1-23767.

Assays for identification of therapeutic targets can be conducted in a variety of ways using methods that are well known to one of ordinary skill in the art. For example, a test

cancerous cell that expresses or overexpresses a differentially expressed gene is contacted with an anti-cancer agent, the effect upon a cancerous phenotype and a biological activity of the candidate gene product assessed. The biological activity of the candidate gene product can be assayed be examining, for example, modulation of expression of a gene encoding the candidate gene product (*e.g.*, as detected by, for example, an increase or decrease in transcript levels or polypeptide levels), or modulation of an enzymatic or other activity of the gene product. The cancerous phenotype can be, for example, cellular proliferation, loss of contact inhibition of growth (*e.g.*, colony formation), tumor growth (*in vitro* or *in vivo*), and the like. Alternatively or in addition, the effect of modulation of a biological activity of the candidate target gene upon cell death/apoptosis or cell cycle regulation can be assessed.

Inhibition or suppression of a cancerous phenotype, or an increase in cell death or apoptosis as a result of modulation of biological activity of a candidate gene product indicates that the candidate gene product is a suitable target for cancer therapy. Assays described infra can be readily adapted for assays for identification of therapeutic targets. Generally such assays are conducted *in vitro*, but many assays can be adapted for *in vivo* analyses, *e.g.*, in an appropriate, art-accepted animal model of the cancer.

<u>Candidate agents</u>

The term "agent" as used herein describes any molecule, *e.g.* protein or pharmaceutical, with the capability of modulating a biological activity of a gene product of a differentially expressed gene. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.* at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including, but not limited to:

peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts (including extracts from human tissue to identify endogenous factors affecting differentially expressed gene products) are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Exemplary candidate agents of particular interest include, but are not limited to, antisense and RNAi polynucleotides, and antibodies, soluble receptors, and the like. Antibodies and soluble receptors are of particular interest as candidate agents where the target differentially expressed gene product is secreted or accessible at the cell-surface (*e.g.*, receptors and other molecule stably-associated with the outer cell membrane).

For method that involve RNAi (RNA interference), a double stranded RNA (dsRNA) molecule is usually used. The dsRNA is prepared to be substantially identical to at least a segment of a subject polynucleotide (e.g. a cDNA or gene). In general, the dsRNA is selected to have at least 70%, 75%, 80%, 85% or 90% sequence identity with the subject polynucleotide over at least a segment of the candidate gene. In other instances, the sequence identity is even higher, such as 95%, 97% or 99%, and in still other instances, there is 100% sequence identity with the subject polynucleotide over at least a segment of the subject polynucleotide. The size of the segment over which there is sequence identity can vary depending upon the size of the subject polynucleotide. In general, however, there is substantial sequence identity over at least 15, 20, 25, 30, 35, 40 or 50 nucleotides. In other instances, there is substantial sequence identity over at least 100, 200, 300, 400, 500 or 1000 nucleotides; in still other instances, there is substantial sequence identity over the entire length of the subject polynucleotide, i.e., the coding and non-coding region of the candidate gene.

Because only substantial sequence similarity between the subject polynucleotide and the dsRNA is necessary, sequence variations between these two species arising from genetic mutations, evolutionary divergence and polymorphisms can be tolerated. Moreover, as described further infra, the dsRNA can include various modified or nucleotide analogs.

Usually the dsRNA consists of two separate complementary RNA strands. However, in some instances, the dsRNA may be formed by a single strand of RNA that is self-complementary, such that the strand loops back upon itself to form a hairpin loop. Regardless of form, RNA duplex formation can occur inside or outside of a cell.

The size of the dsRNA that is utilized varies according to the size of the subject polynucleotide whose expression is to be suppressed and is sufficiently long to be effective in reducing expression of the subject polynucleotide in a cell. Generally, the dsRNA is at least 10-15 nucleotides long. In certain applications, the dsRNA is less than 20, 21, 22, 23, 24 or 25 nucleotides in length. In other instances, the dsRNA is at least 50, 100, 150 or 200 nucleotides in length. The dsRNA can be longer still in certain other applications, such as at least 300, 400, 500 or 600 nucleotides. Typically, the dsRNA is not longer than 3000 nucleotides. The optimal size for any particular subject polynucleotide can be determined by one of ordinary skill in the art without undue experimentation by varying the size of the dsRNA in a systematic fashion and determining whether the size selected is effective in interfering with expression of the subject polynucleotide.

dsRNA can be prepared according to any of a number of methods that are known in the art, including in vitro and in vivo methods, as well as by synthetic chemistry approaches.

In vitro methods. Certain methods generally involve inserting the segment corresponding to the candidate gene that is to be transcribed between a promoter or pair of promoters that are oriented to drive transcription of the inserted segment and then utilizing an appropriate RNA polymerase to carry out transcription. One such arrangement involves positioning a DNA fragment corresponding to the candidate gene or segment thereof into a vector such that it is flanked by two opposable polymerase-specific promoters that can be same or different. Transcription from such promoters produces two complementary RNA strands that can subsequently anneal to form the desired dsRNA. Exemplary plasmids for use in such systems include the plasmid (PCR 4.0 TOPO) (available from Invitrogen). Another example is

the vector pGEM-T (Promega, Madison, WI) in which the oppositely oriented promoters are T7 and SP6; the T3 promoter can also be utilized.

In a second arrangement, DNA fragments corresponding to the segment of the subject polynucleotide that is to be transcribed is inserted both in the sense and antisense orientation downstream of a single promoter. In this system, the sense and antisense fragments are cotranscribed to generate a single RNA strand that is self-complementary and thus can form dsRNA.

Various other *in vitro* methods have been described. Examples of such methods include, but are not limited to, the methods described by Sadher et al. (Biochem. Int. 14:1015, 1987); by Bhattacharyya (Nature 343:484, 1990); and by Livache, et al. (U.S. Patent No. 5,795,715), each of which is incorporated herein by reference in its entirety.

Single-stranded RNA can also be produced using a combination of enzymatic and organic synthesis or by total organic synthesis. The use of synthetic chemical methods enable one to introduce desired modified nucleotides or nucleotide analogs into the dsRNA.

In vivo methods. dsRNA can also be prepared *in vivo* according to a number of established methods (see, e.g., Sambrook, et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed.; Transcription and Translation (B.D. Hames, and S.J. Higgins, Eds., 1984); DNA Cloning, volumes I and II (D.N. Glover, Ed., 1985); and Oligonucleotide Synthesis (M.J. Gait, Ed., 1984, each of which is incorporated herein by reference in its entirety).

Once the single-stranded RNA has been formed, the complementary strands are allowed to anneal to form duplex RNA. Transcripts are typically treated with DNAase and further purified according to established protocols to remove proteins. Usually such purification methods are not conducted with phenol:chloroform. The resulting purified transcripts are subsequently dissolved in RNAase free water or a buffer of suitable composition.

dsRNA is generated by annealing the sense and anti-sense RNA in vitro. Generally, the strands are initially denatured to keep the strands separate and to avoid self-annealing. During the annealing process, typically certain ratios of the sense and antisense strands are combined to facilitate the annealing process. In some instances, a molar ratio of sense to antisense strands of 3:7 is used; in other instances, a ratio of 4:6 is utilized; and in still other instances, the ratio is 1:1.

The buffer composition utilized during the annealing process can in some instances affect the efficacy of the annealing process and subsequent transfection procedure. While some have indicated that the buffered solution used to carry out the annealing process should include a potassium salt such as potassium chloride (e.g. at a concentration of about 80 mM). In some embodiments, the buffer is substantially postassium free. Once single-stranded RNA has annealed to form duplex RNA, typically any single-strand overhangs are removed using an enzyme that specifically cleaves such overhangs (e.g., RNAase A or RNAase T).

Once the dsRNA has been formed, it is introduced into a reference cell, which can include an individual cell or a population of cells (e.g., a tissue, an embryo and an entire organism). The cell can be from essentially any source, including animal, plant, viral, bacterial, fungal and other sources. If a tissue, the tissue can include dividing or nondividing and differentiated or undifferentiated cells. Further, the tissue can include germ line cells and somatic cells. Examples of differentiated cells that can be utilized include, but are not limited to, neurons, glial cells, blood cells, megakaryocytes, lymphocytes, macrophages, neutrophils, eosinophils, basophils, mast cells, leukocytes, granulocytes, keratinocytes, adipocytes, osteoblasts, osteoclasts, hepatocytes, cells of the endocrine or exocrine glands, fibroblasts, myocytes, cardiomyocytes, and endothelial cells. The cell can be an individual cell of an embryo, and can be a blastocyte or an oocyte.

Certain methods are conducted using model systems for particular cellular states (e.g., a disease). For instance, certain methods provided herein are conducted with a cancer cell lines that serves as a model system for investigating genes that are correlated with various cancers.

A number of options can be utilized to deliver the dsRNA into a cell or population of cells such as in a cell culture, tissue or embryo. For instance, RNA can be directly introduced intracellularly. Various physical methods are generally utilized in such instances, such as administration by microinjection (see, e.g., Zernicka-Goetz, et al. (1997) Development 124:1133-1137; and Wianny, et al. (1998) Chromosoma 107: 430-439).

Other options for cellular delivery include permeabilizing the cell membrane and electroporation in the presence of the dsRNA, liposome-mediated transfection, or transfection using chemicals such as calcium phosphate. A number of established gene therapy techniques can also be utilized to introduce the dsRNA into a cell. By introducing a viral construct within a

viral particle, for instance, one can achieve efficient introduction of an expression construct into the cell and transcription of the RNA encoded by the construct.

If the dsRNA is to be introduced into an organism or tissue, gene gun technology is an option that can be employed. This generally involves immobilizing the dsRNA on a gold particle which is subsequently fired into the desired tissue. Research has also shown that mammalian cells have transport mechanisms for taking in dsRNA (see, e.g., Asher, et al. (1969) Nature 223:715-717). Consequently, another delivery option is to administer the dsRNA extracellularly into a body cavity, interstitial space or into the blood system of the mammal for subsequent uptake by such transport processes. The blood and lymph systems and the cerebrospinal fluid are potential sites for injecting dsRNA. Oral, topical, parenteral, rectal and intraperitoneal administration are also possible modes of administration.

The composition introduced can also include various other agents in addition to the dsRNA. Examples of such agents include, but are not limited to, those that stabilize the dsRNA, enhance cellular uptake and/or increase the extent of interference. Typically, the dsRNA is introduced in a buffer that is compatible with the composition of the cell into which the RNA is introduced to prevent the cell from being shocked. The minimum size of the dsRNA that effectively achieves gene silencing can also influence the choice of delivery system and solution composition.

Sufficient dsRNA is introduced into the tissue to cause a detectable change in expression of a taget gene (assuming the candidate gene is in fact being expressed in the cell into which the dsRNA is introduced) using available detection methodologies. Thus, in some instances, sufficient dsRNA is introduced to achieve at least a 5-10% reduction in candidate gene expression as compared to a cell in which the dsRNA is not introduced. In other instances, inhibition is at least 20, 30, 40 or 50%. In still other instances, the inhibition is at least 60, 70, 80, 90 or 95%. Expression in some instances is essentially completely inhibited to undetectable levels.

The amount of dsRNA introduced depends upon various factors such as the mode of administration utilized, the size of the dsRNA, the number of cells into which dsRNA is administered, and the age and size of an animal if dsRNA is introduced into an animal. An appropriate amount can be determined by those of ordinary skill in the art by initially administering dsRNA at several different concentrations for example, for example. In certain

instances when dsRNA is introduced into a cell culture, the amount of dsRNA introduced into the cells varies from about 0.5 to 3 μg per 10^6 cells.

A number of options are available to detect interference of candidate gene expression (i.e., to detect candidate gene silencing). In general, inhibition in expression is detected by detecting a decrease in the level of the protein encoded by the candidate gene, determining the level of mRNA transcribed from the gene and/or detecting a change in phenotype associated with candidate gene expression.

USE OF POLYPEPTIDES TO SCREEN FOR PEPTIDE ANALOGS AND ANTAGONISTS

Polypeptides encoded by differentially expressed genes identified herein can be used to screen peptide libraries to identify binding partners, such as receptors, from among the encoded polypeptides. Peptide libraries can be synthesized according to methods known in the art (*see, e.g.,* USPN 5,010,175 and WO 91/17823).

Agonists or antagonists of the polypeptides of the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, etc. The assay conditions ideally should resemble the conditions under which the native activity is exhibited *in vivo*, that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

Such screening and experimentation can lead to identification of a polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide described herein, and at least one peptide agonist or antagonist of the binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

VACCINES AND USES

The differentially expressed nucleic acids and polypeptides produced by the nucleic acids of the invention can also be used to modulate primary immune response to prevent or treat cancer. Every immune response is a complex and intricately regulated sequence of events involving several cell types. It is triggered when an antigen enters the body and encounters a specialized class of cells called antigen-presenting cells (APCs). These APCs capture a minute amount of the antigen and display it in a form that can be recognized by antigen-specific helper T lymphocytes. The helper (Th) cells become activated and, in turn, promote the activation of other classes of lymphocytes, such as B cells or cytotoxic T cells. The activated lymphocytes then proliferate and carry out their specific effector functions, which in many cases successfully activate or eliminate the antigen. Thus, activating the immune response to a particular antigen associated with a cancer cell can protect the patient from developing cancer or result in lymphocytes eliminating cancer cells expressing the antigen.

Gene products, including polypeptides, mRNA (particularly mRNAs having distinct secondary and/or tertiary structures), cDNA, or complete gene, can be prepared and used in vaccines for the treatment or prevention of hyperproliferative disorders and cancers. The nucleic acids and polypeptides can be utilized to enhance the immune response, prevent tumor progression, prevent hyperproliferative cell growth, and the like. Methods for selecting nucleic acids and polypeptides that are capable of enhancing the immune response are known in the art. Preferably, the gene products for use in a vaccine are gene products which are present on the surface of a cell and are recognizable by lymphocytes and antibodies.

The gene products may be formulated with pharmaceutically acceptable carriers into pharmaceutical compositions by methods known in the art. The composition is useful as a vaccine to prevent or treat cancer. The composition may further comprise at least one co-immunostimulatory molecule, including but not limited to one or more major histocompatibility complex (MHC) molecules, such as a class I or class II molecule, preferably a class I molecule. The composition may further comprise other stimulator molecules including B7.1, B7.2, ICAM-1, ICAM-2, LFA-1, LFA-3, CD72 and the like, immunostimulatory polynucleotides (which comprise an 5'-CG-3' wherein the cytosine is unmethylated), and cytokines which include but are not limited to IL-1 through IL-15, TNF-α, IFN-γ, RANTES, G-CSF, M-CSF, IFN-α, CTAP

III, ENA-78, GRO, I-309, PF-4, IP-10, LD-78, MGSA, MIP-1α, MIP-1β, or combination thereof, and the like for immunopotentiation. In one embodiment, the immunopotentiators of particular interest are those that facilitate a Th1 immune response.

The gene products may also be prepared with a carrier that will protect the gene products against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, polylactic acid, and the like. Methods for preparation of such formulations are known in the art.

In the methods of preventing or treating cancer, the gene products may be administered via one of several routes including but not limited to transdermal, transmucosal, intravenous, intramuscular, subcutaneous, intradermal, intraperitoneal, intrathecal, intrapleural, intrauterine, rectal, vaginal, topical, intratumor, and the like. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be by nasal sprays or suppositories. For oral administration, the gene products are formulated into conventional oral administration form such as capsules, tablets, elixirs and the like.

The gene product is administered to a patient in an amount effective to prevent or treat cancer. In general, it is desirable to provide the patient with a dosage of gene product of at least about 1 pg per Kg body weight, preferably at least about 1 ng per Kg body weight, more preferably at least about 1 μg or greater per Kg body weight of the recipient. A range of from about 1 ng per Kg body weight to about 100 mg per Kg body weight is preferred although a lower or higher dose may be administered. The dose is effective to prime, stimulate and/or cause the clonal expansion of antigen-specific T lymphocytes, preferably cytotoxic T lymphocytes, which in turn are capable of preventing or treating cancer in the recipient. The dose is administered at least once and may be provided as a bolus or a continuous administration. Multiple administrations of the dose over a period of several weeks to months may be preferable. Subsequent doses may be administered as indicated.

In another method of treatment, autologous cytotoxic lymphocytes or tumor infiltrating lymphocytes may be obtained from a patient with cancer. The lymphocytes are grown in

culture, and antigen-specific lymphocytes are expanded by culturing in the presence of the specific gene products alone or in combination with at least one co-immunostimulatory molecule with cytokines. The antigen-specific lymphocytes are then infused back into the patient in an amount effective to reduce or eliminate the tumors in the patient. Cancer vaccines and their uses are further described in USPN 5,961,978; USPN 5,993,829; USPN 6,132,980; and WO 00/38706.

PHARMACEUTICAL COMPOSITIONS AND USES

Pharmaceutical compositions can comprise polypeptides, receptors that specifically bind a polypeptide produced by a differentially expressed gene (*e.g.*, antibodies, or polynucleotides (including antisense nucleotides and ribozymes) of the claimed invention in a therapeutically effective amount. The compositions can be used to treat primary tumors as well as metastases of primary tumors. In addition, the pharmaceutical compositions can be used in conjunction with conventional methods of cancer treatment, *e.g.*, to sensitize tumors to radiation or conventional chemotherapy.

Where the pharmaceutical composition comprises a receptor (such as an antibody) that specifically binds to a gene product encoded by a differentially expressed gene, the receptor can be coupled to a drug for delivery to a treatment site or coupled to a detectable label to facilitate imaging of a site comprising cancer cells. Methods for coupling antibodies to drugs and detectable labels are well known in the art, as are methods for imaging using detectable labels.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature.

The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present

invention, an effective dose will generally be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles.

Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, e.g., mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in *Remington: The Science and Practice of Pharmacy* (1995) Alfonso Gennaro, Lippincott, Williams, & Wilkins.

DELIVERY METHODS

Once formulated, the compositions contemplated by the invention can be (1) administered directly to the subject (*e.g.*, as polynucleotide, polypeptides, small molecule agonists or antagonists, and the like); or (2) delivered ex vivo, to cells derived from the subject (*e.g.*, as in *ex vivo* gene therapy). Direct delivery of the compositions will generally be accomplished by parenteral injection, e.g., subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumoral or to the interstitial space of a tissue. Other modes of

administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.

Methods for the ex vivo delivery and reimplantation of transformed cells into a subject are known in the art and described in *e.g.*, International Publication No. WO 93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both ex vivo and in vitro applications can be accomplished by, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Once differential expression of a gene corresponding to a polynucleotide described herein has been found to correlate with a proliferative disorder, such as neoplasia, dysplasia, and hyperplasia, the disorder can be amenable to treatment by administration of a therapeutic agent based on the provided polynucleotide, corresponding polypeptide or other corresponding molecule (e.g., antisense, ribozyme, etc.). In other embodiments, the disorder can be amenable to treatment by administration of a small molecule drug that, for example, serves as an inhibitor (antagonist) of the function of the encoded gene product of a gene having increased expression in cancerous cells relative to normal cells or as an agonist for gene products that are decreased in expression in cancerous cells (*e.g.*, to promote the activity of gene products that act as tumor suppressors).

The dose and the means of administration of the inventive pharmaceutical compositions are determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. For example, administration of polynucleotide therapeutic composition agents includes local or systemic administration, including injection, oral administration, particle gun or catheterized administration, and topical administration. In general, the therapeutic polynucleotide composition contains an expression construct comprising a promoter operably linked to a polynucleotide of at least 12, 22, 25, 30, or 35 contiguous nt of the polynucleotide disclosed herein. Various methods can be used to administer the therapeutic composition directly to a specific site in the body. For example, a small metastatic lesion is located and the therapeutic

composition injected several times in several different locations within the body of the tumor. Alternatively, arteries which serve a tumor are identified, and the therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor. A tumor that has a necrotic center is aspirated and the composition injected directly into the now empty center of the tumor. The antisense composition is directly administered to the surface of the tumor, for example, by topical application of the composition. X-ray imaging is used to assist in certain of the above delivery methods.

Targeted delivery of therapeutic compositions containing an antisense polynucleotide, subgenomic polynucleotides, or antibodies to specific tissues can also be used. Receptor-mediated DNA delivery techniques are described in, for example, Findeis et al., Trends Biotechnol. (1993) 11:202; Chiou et al., Gene Therapeutics: Methods And Applications Of Direct Gene Transfer (J.A. Wolff, ed.) (1994); Wu et al., J. Biol. Chem. (1988) 263:621; Wu et al., J. Biol. Chem. (1994) 269:542; Zenke et al., Proc. Natl. Acad. Sci. (USA) (1990) 87:3655; Wu et al., J. Biol. Chem. (1991) 266:338. Therapeutic compositions containing a polynucleotide are administered in a range of about 100 ng to about 200 mg of DNA for local administration in a gene therapy protocol. Concentration ranges of about 500 ng to about 50 mg, about 1 μg to about 2 mg, about 5 μg to about 500 μg, and about 20 μg to about 100 :g of DNA can also be used during a gene therapy protocol. Factors such as method of action (e.g., for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations that will affect the dosage required for ultimate efficacy of the antisense subgenomic polynucleotides.

The therapeutic polynucleotides and polypeptides of the present invention can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally, Jolly, Cancer Gene Therapy (1994) 1:51; Kimura, Human Gene Therapy (1994) 5:845; Connelly, Human Gene Therapy (1995) 1:185; and Kaplitt, Nature Genetics (1994) 6:148). Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

Viral-based vectors for delivery of a desired polynucleotide and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (see, e.g., WO 90/07936; WO 94/03622; WO 93/25698; WO

93/25234; USPN 5, 219,740; WO 93/11230; WO 93/10218; USPN 4,777,127; GB Patent No. 2,200,651; EP 0 345 242; and WO 91/02805), alphavirus-based vectors (e.g., Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), and adeno-associated virus (AAV) vectors (see, e.g., WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655). Administration of DNA linked to killed adenovirus as described in Curiel, Hum. Gene Ther. (1992) 3:147 can also be employed.

Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone (see, e.g., Curiel, Hum. Gene Ther. (1992) 3:147); ligand-linked DNA (see, e.g., Wu, J. Biol. Chem. (1989) 264:16985); eukaryotic cell delivery vehicles cells (see, e.g., USPN 5,814,482; WO 95/07994; WO 96/17072; WO 95/30763; and WO 97/42338) and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and USPN 5,580,859. Liposomes that can act as gene delivery vehicles are described in USPN 5,422,120; WO 95/13796; WO 94/23697; WO 91/14445; and EP 0524968. Additional approaches are described in Philip, *Mol. Cell Biol.* (1994) *14*:2411, and in Woffendin, *Proc. Natl. Acad. Sci.* (1994) *91*:1581.

The sequences disclosed in this patent application were disclosed in several earlier patent applications. The relationship between the SEQ ID NOS in those earlier application and the SEQ ID NOS disclosed herein is shown in Tables 161 and 162.

Table 161: relationship between SEQ ID NOs. this patent application and SEQ ID NOs of parent patent applications

parent case	parent application no.	filing date	SEQ IDs in parent case	corresponding SEQ IDs in this patent application
1480	10/076,555	February 15, 2002	1-844	1-844
1481	09/297,648	March 10, 2000	1-5252	845-6096
1487	09/313,292	May 13, 1999	1-2707	6097-8803
1490	09/854,124	May 10, 2001	1-37	8804-8840

1492	09/404,706	September 23, 1999	1-1079	8841-9919
1598	10/629,771	July 28, 2003	1-3351	9920-13270
1624	09/803,719	March 9, 2001	1-2396	13271-15666
1625	10/609,021	June 26, 2003	1-324	15667-15990
15990	10/615,618	July 7, 2003	1-6010	15991-22000
16252	10/012,697	December 7, 2001	1-1568	22001-23568
18790	60/532,830	December 23, 2003	1-199	23569-23767

The disclosures of all prior U.S. applications to which the present application claims priority, which includes those U.S. applications referenced in the table above as well as their respective priority applications, are each incorporated herein by referenced in their entireties for all purposes, including the disclosures found in the Sequence Listings, tables, figures and Examples.

EXAMPLES

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Example 1: Source of Biological Materials and Overview of Novel Polynucleotides Expressed by the Biological Materials

Human colon cancer cell line Km12L4-A (Morika, W. A. K. et al., *Cancer Research* (1988) *48*:6863) was used to construct a cDNA library from mRNA isolated from the cells. As described in the above overview, a total of 4,693 sequences expressed by the Km12L4-A cell line were isolated and analyzed; most sequences were about 275-300 nucleotides in length. The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B_2 surgical specimen (Morikawa *et al. Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman *et al. Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling *et al. Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as a model cell line for the study of colon cancer (see, *e.g.*, Moriakawa *et al.*, *supra*; Radinsky *et al. Clin. Cancer Res.* (1995) *1*:19; Yeatman *et al.*, (1995) *supra*; Yeatman *et al. Clin. Exp. Metastasis* (1996) *14*:246).

The sequences were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol.* 266:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis Techniques" Adams *et al.*, eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie *et al. Comput. Chem.* (1993) 17:191). Generally, masking does not influence the final search results, except to eliminate of relative little interest due to their lox complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats. Masking resulted in the elimination of 43 sequences. The remaining sequences were then used in a BLASTN vs. Genbank search with search parameters of greater than 70% overlap, 99% identity, and a p value of less than 1×10^{-40}, which search resulted in the discarding of 1,432 sequences. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database search: (1) unknown (no hits in the Genbank search), (2) weak similarity (greater than 45% identity and p value of less than 1 x 10^{-5}), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1 x 10^{-5}). This search resulted in discard of 98 sequences as having greater than 70% overlap, greater than 99% identity, and p value of less than 1 x 10^{-40}.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search resulted in discard of 1771 sequences (sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1 x 10^{-40}; sequences with a p value of less than 1 x 10^{-65} when compared to a database sequence of human origin were also excluded). Second, a BLASTN vs. Patent GeneSeq database resulted in discard of 15 sequences (greater than 99% identity; p value less than 1 x 10^{-40}; greater than 99% overlap).

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1 x 10^{-111} in relation to a database sequence of human origin were specifically excluded. The final result provided the 404 sequences listed in the accompanying Sequence Listing. The Sequence Listing is arranged beginning with sequences with no similarity to any sequence in a database searched, and ending with sequences with the greatest similarity. Each identified polynucleotide represents sequence from at least a partial mRNA transcript. Polynucleotides that were determined to be novel were assigned a sequence identification number.

The novel polynucleotides and were assigned sequence identification numbers SEQ ID NOS: 1-404. The DNA sequences corresponding to the novel polynucleotides are provided in the Sequence Listing. The majority of the sequences are presented in the Sequence Listing in the 5' to 3' direction. A small number, 25, are listed in the Sequence Listing in the 5' to 3' direction but the sequence as written is actually 3' to 5'. These sequences are readily identified with the designation "AR" in the Sequence Name in Table 1 (inserted before the claims). The sequences correctly listed in the 5' to 3' direction in the Sequence Listing are designated "AF." The Sequence Listing filed herewith therefore contains 25 sequences listed in the reverse order,

namely SEQ ID NOS:47, 97, 137, 171, 173, 179, 182, 194, 200, 202, 213, 227, 258, 264, 275, 302, 313, 324, 329, 330, 331, 338, 358, 379, and 404.

Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOS: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene.

In order to confirm the sequences of SEQ ID NOS:1-404, inserts of the clones corresponding to these polynucleotides were re-sequenced. These "validation" sequences are provided in SEQ ID NOS:405-800. These validation sequences were often longer than the original polynucleotide sequences. They validate, and thus often provide additional sequence information. Validation sequences can be correlated with the original sequences they validate by identifying those sequences of SEQ ID NOS:1-404 and the validation sequences of SEQ ID NOS:405-800 that share the same clone name in Table 1.

Table 1. Sequence identification numbers, cluster ID, sequence name, and clone name

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
1	4635	RTA00000180AF.i.20.1	M00001429B:A11
2		RTA00000185AF.n.12.1	M00001608D:A11
3	4622	RTA00000187AF.m.15.2	M00001686A:E06
4	3706	RTA00000191AF.i.17.2	M00004068B:A01
5	36535	RTA00000181AF.f.5.1	M00001449A:G10
6	3990	RTA00000183AF.j.11.1	M00001532B:A06
7	5319	RTA00000192AF.i.12.1	M00004169C:C12
8	36393	RTA00000180AF.c.2.1	M00001417A:E02
9	2623	RTA00000183AF.a.6.1	M00001497A:G02
10	7587	RTA00000178AF.n.24.1	M00001387B:G03
11	7065	RTA00000137A.g.6.1	M00001557A:D02
12	10539	RTA00000187AF.l.7.1	M00001680D:F08
13	27250	RTA00000181AF.g.10.1	M00001450A:D08
14	5556	RTA00000179AF.n.10.1	M00001407B:D11
15		RTA00000192AF.m.12.1	M00004191D:B11
16	8761	RTA00000184AF.k.12.1	M00001557D:D09
17	4622	RTA00000189AF.g.1.1	M00003856B:C02

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
18	11460	RTA00000187AF.g.12.1	M00001676B:F05
19	16283	RTA00000120A.o.20.1	M00001467A:D08
20	3430	RTA00000191AF.a.9.1	M00003981A:E10
21	7065	RTA00000184AF.j.21.1	M00001557A:D02
22		RTA00000182AF.l.20.1	M00001488B:F12
23		RTA00000123A.g.19.1	M00001531A:H11
24	16918	RTA00000193AF.a.16.1	M00004223A:G10
25	16914	RTA00000193AF.f.5.1	M00004275C:C11
26	40108	RTA00000187AF.o.24.1	M00003741D:C09
27	14286	RTA00000193AF.f.22.1	M00004283B:A04
28	17004	RTA00000186AF.b.21.1	M00001617C:E02
29		RTA00000180AF.g.22.1	M00001426B:D12
30	13272	RTA00000192AF.e.3.1	M00004138B:H02
31		RTA00000194AF.f.4.1	M00005180C:G03
32	32663	RTA00000118A.l.8.1	M00001450A:A11
33		RTA00000180AF.a.9.1	M00001414A:B01
34	5832	RTA00000178AF.o.23.1	M00001388D:G05
35	7801	RTA00000181AF.c.21.1	M00001446A:F05
36	76760	RTA00000187AF.a.15.1	M00001657D:F08
37	40132	RTA00000178AF.c.7.1	M00001365C:C10
38		RTA00000183AF.e.1.1	M00001505C:C05
39	4016	RTA00000118A.c.4.1	M00001395A:C03
40	5382	RTA00000187AF.m.23.2	M00001688C:F09
41	5693	RTA00000190AF.p.17.2	M00003978B:G05
42	307	RTA00000136A.o.4.2	M00001552A:B12
43	39833	RTA00000178AF.i.23.1	M00001378B:B02
44		RTA00000193AF.m.5.1	M00004359B:G02
45	5325	RTA00000191AF.o.6.1	M00004093D:B12
46	5325	RTA00000191AF.o.6.2	M00004093D:B12
47	18957	RTA00000190AR.m.9.1	M00003958A:H02
48	39508	RTA00000120A.o.2.1	M00001467A:D04
49	22390	RTA00000136A.j.13.1	M00001551A:G06
50	12170	RTA00000125A.h.18.4	M00001544A:E03
51	4393	RTA00000187AF.n.17.1	M00001693C:G01
52	19	RTA00000182AF.b.7.1	M00001463C:B11
53		RTA00000193AF.c.21.1	M00004249D:F10
54	7899	RTA00000189AF.c.10.1	M00003837D:A01

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
55	40073	RTA00000191AF.e.3.1	M00004028D:C05
56	7005	RTA00000179AF.o.22.1	M00001410A:D07
57		RTA00000187AF.h.22.1	M00001679A:F06
58	18957	RTA00000190AF.m.9.2	M00003958A:H02
59	18957	RTA00000183AF.h.23.1	M00001528A:F09
60	16283	RTA00000182AF.c.22.1	M00001467A:D08
61	6974	RTA00000183AF.d.9.1	M00001504C:H06
62	2623	RTA00000183AF.b.14.1	M00001500A:E11
63	9105	RTA00000191AF.a.21.2	M00003983A:A05
64	13238	RTA00000181AF.m.4.1	M00001455A:E09
65	5749	RTA00000185AF.a.19.1	M00001571C:H06
66	6455	RTA00000193AF.b.9.1	M00004229B:F08
67	23001	RTA00000185AF.c.24.1	M00001578B:E04
68	6455	RTA00000192AF.g.23.1	M00004157C:A09
69	13595	RTA00000189AF.f.8.1	M00003851B:D10
70	39442	RTA00000120A.o.21.1	M00001467A:E10
71	17036	RTA00000191AF.f.13.1	M00004035D:B06
72		RTA00000183AF.g.9.1	M00001513B:G03
73	7005	RTA00000181AF.k.24.1	M00001454B:C12
74	6268	RTA00000126A.o.23.1	M00001551A:B10
75	16130	RTA00000119A.c.13.1	M00001453A:E11
76	23201	RTA00000187AF.a.14.1	M00001657D:C03
77	5321	RTA00000183AF.k.8.1	M00001534A:F09
78	13157	RTA00000186AF.a.6.1	M00001614C:F10
79	2102	RTA00000193AF.n.7.1	M00004377C:F05
80	1058	RTA00000126A.e.20.3	M00001548A:H09
81	40392	RTA00000180AF.j.8.1	M00001429D:D07
82		RTA00000183AF.e.23.1	M00001506D:A09
83	11476	RTA00000187AF.p.19.1	M00003747D:C05
84	3584	RTA00000177AF.h.20.1	M00001349B:B08
85	10470	RTA00000180AF.f.18.1	M00001424B:G09
86	39425	RTA00000133A.f.1.1	M00001470A:C04
87	5175	RTA00000184AF.f.3.1	M00001550A:G01
88	13576	RTA00000189AF.o.13.1	M00003885C:A02
89	7665	RTA00000134A.l.19.1	M00001535A:B01
90	16927	RTA00000177AF.h.9.3	M00001348B:B04
91	6660	RTA00000187AF.h.15.1	M00001679A:A06

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
92	2433	RTA00000191AF.a.15.2	M00003982C:C02
93	5097	RTA00000134A.k.1.1	M00001534A:D09
94	21847	RTA00000193AF.j.9.1	M00004318C:D10
95	3277	RTA00000138A.l.5.1	M00001624A:B06
96	5708	RTA00000184AF.g.12.1	M00001552B:D04
97	945	RTA00000178AR.a.20.1	M00001362C:H11
98	16269	RTA00000178AF.p.1.1	M00001389A:C08
99		RTA00000183AF.c.24.1	M00001504A:E01
100	16731	RTA00000181AF.a.20.1	M00001442C:D07
101	12439	RTA00000190AF.o.24.1	M00003975A:G11
102	3162	RTA00000177AF.j.12.3	M00001351B:A08
103		RTA00000194AF.b.19.1	M00004505D:F08
104		RTA00000193AF.n.15.1	M00004384C:D02
105		RTA00000186AF.n.7.1	M00001651A:H01
106	10717	RTA00000181AF.d.10.1	M00001447A:G03
107	4573	RTA00000189AF.j.12.1	M00003871C:E02
108		RTA00000186AF.h.14.1	M00001632D:H07
109	11443	RTA00000192AF.l.13.2	M00004185C:C03
110	5892	RTA00000184AF.d.11.1	M00001548A:E10
111	3162	RTA00000177AF.j.12.1	M00001351B:A08
112	10470	RTA00000185AF.k.6.1	M00001597D:C05
113	17055	RTA00000187AF.m.3.1	M00001682C:B12
114	2030	RTA00000193AF.m.20.1	M00004372A:A03
115	6558	RTA00000184AF.m.21.1	M00001560D:F10
116	23255	RTA00000190AF.j.4.1	M00003922A:E06
117	9577	RTA00000179AF.o.17.1	M00001409C:D12
118		RTA00000180AF.a.11.1	M00001414C:A07
119	8	RTA00000181AF.e.17.1	M00001448D:C09
120	67907	RTA00000188AF.g.11.1	M00003774C:A03
121	12081	RTA00000133A.d.14.2	M00001469A:C10
122	2448	RTA00000119A.j.21.1	M00001460A:F06
123	3389	RTA00000189AF.g.3.1	M00003857A:G10
124	39174	RTA00000124A.n.13.1	M00001541A:H03
125	24488	RTA00000190AF.n.16.1	M00003968B:F06
126	8210	RTA00000192AF.n.13.1	M00004197D:H01
127		RTA00000135A.l.2.2	M00001545A:B02
128	40455	RTA00000190AF.m.10.2	M00003958C:G10

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
129	9577	RTA00000180AF.d.23.1	M00001421C:F01
130	13183	RTA00000192AF.a.24.1	M00004114C:F11
131	5214	RTA00000186AF.g.11.1	M00001630B:H09
132	67252	RTA00000187AF.o.6.1	M00001716D:H05
133	3108	RTA00000188AF.d.24.1	M00003763A:F06
134	2464	RTA00000178AF.n.18.1	M00001387A:C05
135	36313	RTA00000181AF.e.23.1	M00001448D:H01
136	23255	RTA00000177AF.e.14.3	M00001343D:H07
137	7985	RTA00000182AR.j.2.1	M00001481D:A05
138	8286	RTA00000183AF.o.1.1	M00001540A:D06
139	22195	RTA00000180AF.g.7.1	M00001425B:H08
140	4573	RTA00000184AF.h.9.1	M00001553B:F12
141	26875	RTA00000187AF.i.1.1	M00001679A:F10
142	7187	RTA00000177AF.i.8.2	M00001350A:H01
143	86859	RTA00000118A.p.8.1	M00001452A:B12
144	4623	RTA00000185AF.f.4.1	M00001586C:C05
145		RTA00000121A.c.10.1	M00001469A:A01
146	10185	RTA00000183AF.d.5.1	M00001504C:A07
147		RTA00000183AF.p.4.1	M00001542B:B01
148	15069	RTA00000191AF.l.6.1	M00004081C:D10
149	39304	RTA00000118A.j.21.1	M00001450A:A02
150	8672	RTA00000190AF.f.11.1	M00003909D:C03
151	13576	RTA00000177AF.g.16.1	M00001347A:B10
152	6293	RTA00000185AF.e.11.1	M00001583D:A10
153	16977	RTA00000192AF.g.3.1	M00004151D:B08
154	5345	RTA00000189AF.l.19.1	M00003879B:C11
155	4905	RTA00000193AF.e.14.1	M00004269D:D06
156	17036	RTA00000191AF.j.10.1	M00004072B:B05
157	5417	RTA00000191AF.h.19.1	M00004059A:D06
158	7172	RTA00000178AF.f.9.1	M00001371C:E09
159	40044	RTA00000186AF.d.1.1	M00001621C:C08
160	4386	RTA00000184AF.j.4.1	M00001556B:C08
161	40044	RTA00000183AF.g.22.1	M00001514C:D11
162	9685	RTA00000183AF.c.11.1	M00001501D:C02
163	22155	RTA00000185AF.n.9.1	M00001608B:E03
164	10515	RTA00000189AF.f.18.1	M00003853A:F12
165	6539	RTA00000185AF.d.11.1	M00001579D:C03

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
166	15066	RTA00000180AF.e.24.1	M00001423B:E07
167	4261	RTA00000180AF.h.5.1	M00001426D:C08
168	13864	RTA00000125A.m.9.1	M00001545A:D08
169	6539	RTA00000189AF.d.22.1	M00003844C:B11
170	11465	RTA00000185AF.m.19.1	M00001607A:E11
171	3266	RTA00000184AR.g.1.1	M00001551C:G09
172	102	RTA00000184AF.o.5.1	M00001563B:F06
173	16970	RTA00000181AR.i.18.2	M00001452C:B06
174	12971	RTA00000193AF.a.20.1	M00004223D:E04
175	5007	RTA00000177AF.g.2.1	M00001346A:F09
176	3765	RTA00000135A.d.1.1	M00001541A:D02
177	11294	RTA00000184AF.j.6.1	M00001556B:G02
178	3681	RTA00000131A.g.15.2	M00001449A:D12
179	9283	RTA00000181AR.m.21.2	M00001455D:F09
180	18699	RTA00000182AF.m.16.1	M00001490B:C04
181	86110	RTA00000181AF.f.12.1	M00001449C:D06
182	39648	RTA00000178AR.l.8.2	M00001383A:C03
183	7337	RTA00000123A.b.17.1	M00001528A:C04
184	1334	RTA00000178AF.j.7.1	M00001379A:A05
185	17076	RTA00000188AF.d.21.1	M00003762C:B08
186	22794	RTA00000138A.b.5.1	M00001601A:D08
187	39171	RTA00000186AF.l.7.1	M00001644C:B07
188	8551	RTA00000179AF.p.21.1	M00001412B:B10
189	5857	RTA00000118A.g.14.1	M00001449A:A12
190	9443	RTA00000183AF.c.1.1	M00001500C:E04
191	9457	RTA00000193AF.i.14.2	M00004307C:A06
192	7206	RTA00000182AF.o.15.1	M00001494D:F06
193	22979	RTA00000178AF.k.22.1	M00001382C:A02
194	40455	RTA00000190AR.m.10.1	M00003958C:G10
195	7221	RTA00000191AF.p.9.1	M00004105C:A04
196		RTA00000191AF.j.9.1	M00004072A:C03
197	7239	RTA00000126A.m.4.2	M00001550A:A03
198	31587	RTA00000189AF.l.20.1	M00003879B:D10
199	16317	RTA00000190AF.e.6.1	M00003907D:H04
200	13576	RTA00000189AR.o.13.1	M00003885C:A02
201	5779	RTA00000177AF.g.14.3	M00001346D:G06
202	6124	RTA00000191AR.e.2.3	M00004028D:A06

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
203	9952	RTA00000180AF.c.20.1	M00001418B:F03
204		RTA00000188AF.i.8.1	M00003784D:D12
205	5779	RTA00000177AF.g.14.1	M00001346D:G06
206	39490	RTA00000128A.b.4.1	M00001557A:F03
207	4416	RTA00000187AF.h.13.1	M00001678D:F12
208	4009	RTA00000179AF.e.20.1	M00001396A:C03
209	5336	RTA00000183AF.b.13.1	M00001500A:C05
210	39186	RTA00000121A.p.15.1	M00001512A:A09
211	40122	RTA00000190AF.n.23.1	M00003970C:B09
212	12532	RTA00000190AF.g.2.1	M00003912B:D01
213	8078	RTA00000177AR.l.13.1	M00001353A:G12
214	3900	RTA00000190AF.g.13.1	M00003914C:F05
215	7589	RTA00000120A.p.23.1	M00001468A:F05
216	8298	RTA00000127A.d.19.1	M00001553A:H06
217	4443	RTA00000177AF.b.20.4	M00001341A:E12
218	26295	RTA00000193AF.i.24.2	M00004312A:G03
219	3389	RTA00000183AF.m.19.1	M00001537B:G07
220	7015	RTA00000187AF.f.18.1	M00001673C:H02
221	8526	RTA00000180AF.d.1.1	M00001418D:B06
222	4665	RTA00000186AF.m.3.1	M00001648C:A01
223	1399	RTA00000129A.o.10.1	M00001604A:B10
224	9244	RTA00000127A.l.3.1	M00001556A:C09
225		RTA00000179AF.j.13.1	M00001400B:H06
226	82498	RTA00000118A.m.10.1	M00001450A:B12
227	35702	RTA00000187AR.c.15.2	M00001663A:E04
228	38759	RTA00000120A.m.12.3	M00001467A:B07
229	39648	RTA00000178AF.l.8.1	M00001383A:C03
230	19105	RTA00000133A.e.15.1	M00001469A:H12
231	85064	RTA00000131A.m.23.1	M00001452A:F05
232	9285	RTA00000191AF.m.18.1	M00004086D:G06
233	9285	RTA00000190AF.d.7.1	M00003906C:E10
234	39391	RTA00000138A.c.3.1	M00001604A:F05
235		RTA00000178AF.d.20.1	M00001368D:E03
236	39498	RTA00000119A.j.20.1	M00001460A:F12
237	7798	RTA00000189AF.k.12.1	M00003876D:E12
238	7798	RTA00000189AF.c.18.1	M00003839A:D08
239	19829	RTA00000125A.h.24.4	M00001544A:G02

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
240		RTA00000188AF.d.11.1	M00003761D:A09
241	4275	RTA00000120A.j.14.1	M00001466A:E07
242	22113	RTA00000125A.c.7.1	M00001542A:A09
243	40314	RTA00000186AF.c.15.1	M00001619C:F12
244	10944	RTA00000126A.h.17.2	M00001549A:D08
245	39809	RTA00000190AF.e.3.1	M00003907D:A09
246	22085	RTA00000135A.e.5.2	M00001541A:F07
247	19255	RTA00000135A.m.18.1	M00001545A:C03
248	14311	RTA00000192AF.o.2.1	M00004203B:C12
249	8479	RTA00000189AF.j.22.1	M00003875C:G07
250		RTA00000189AF.j.23.1	M00003875D:D11
251	4193	RTA00000184AF.e.13.1	M00001549B:F06
252	22814	RTA00000184AF.h.14.1	M00001553D:D10
253	39563	RTA00000179AF.k.20.1	M00001402A:E08
254	39420	RTA00000134A.o.23.1	M00001537A:F12
255	11589	RTA00000177AF.b.17.4	M00001340D:F10
256	4937	RTA00000191AF.p.21.1	M00004108A:E06
257	39412	RTA00000133A.k.17.1	M00001511A:H06
258	4837	RTA00000185AR.k.3.2	M00001597C:H02
259	13046	RTA00000193AF.h.19.1	M00004296C:H07
260	4141	RTA00000177AF.p.20.3	M00001361A:A05
261	38085	RTA00000123A.e.15.1	M00001531A:D01
262		RTA00000189AF.p.8.1	M00003891C:H09
263	11451	RTA00000192AF.p.17.1	M00004214C:H05
264	14507	RTA00000189AR.l.23.2	M00003879D:A02
265	40054	RTA00000180AF.p.10.1	M00001439C:F08
266	39423	RTA00000134A.k.22.1	M00001535A:F10
267	39453	RTA00000135A.g.11.1	M00001542A:E06
268	10751	RTA00000187AF.k.7.1	M00001679D:D03
269	10751	RTA00000187AF.k.6.1	M00001679D:D03
270	78091	RTA00000187AF.j.6.1	M00001679C:F01
271	39539	RTA00000127A.i.21.1	M00001555A:B02
272		RTA00000182AF.l.15.1	M00001487B:H06
273		RTA00000194AF.d.13.1	M00004896A:C07
274		RTA00000128A.c.20.1	M00001558A:H05
275	9283	RTA00000181AR.m.22.2	M00001455D:F09
276	39168	RTA00000121A.l.10.1	M00001507A:H05

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
277	39458	RTA00000126A.p.15.2	M00001552A:D11
278	14391	RTA00000177AF.m.17.3	M00001355B:G10
279	39195	RTA00000137A.c.16.1	M00001555A:C01
280	7212	RTA00000193AF.b.14.1	M00004230B:C07
281	4015	RTA00000136A.e.12.1	M00001549A:B02
282	12977	RTA00000189AF.j.19.1	M00003875B:F04
283		RTA00000178AF.m.13.1	M00001384B:A11
284	14391	RTA00000191AF.l.7.1	M00004081C:D12
285		RTA00000194AF.c.23.1	M00004691D:A05
286		RTA00000181AF.b.7.1	M00001443B:F01
287	8358	RTA00000183AF.i.5.1	M00001528B:H04
288	1267	RTA00000125A.o.5.1	M00001546A:G11
289		RTA00000189AF.f.7.1	M00003851B:D08
290	16347	RTA00000184AF.e.15.1	M00001549C:E06
291	7899	RTA00000193AF.a.17.1	M00004223B:D09
292	2379	RTA00000178AF.a.6.1	M00001361D:F08
293	39478	RTA00000133A.i.5.1	M00001471A:B01
294	39392	RTA00000134A.m.16.1	M00001536A:C08
295	5053	RTA00000184AF.o.12.1	M00001564A:B12
296	16999	RTA00000185AF.k.9.1	M00001598A:G03
297	39180	RTA00000126A.n.8.2	M00001551A:F05
298	1037	RTA00000121A.f.8.1	M00001470A:B10
299	6867	RTA00000178AF.e.12.1	M00001370A:C09
300	10539	RTA00000183AF.a.24.1	M00001499B:A11
301	41633	RTA00000118A.g.16.1	M00001449A:B12
302	23218	RTA00000187AR.c.5.2	M00001662C:A09
303	39380	RTA00000129A.e.24.1	M00001587A:B11
304		RTA00000185AF.d.24.1	M00001582D:F05
305		RTA00000177AF.o.4.3	M00001358C:C06
306	6974	RTA00000184AF.a.15.1	M00001544B:B07
307		RTA00000185AF.g.11.1	M00001590B:F03
308	15855	RTA00000184AF.j.1.1	M00001556A:H01
309	84328	RTA00000118A.p.10.1	M00001452A:B04
310	10145	RTA00000120A.g.12.1	M00001465A:B11
311	39805	RTA00000177AF.c.21.3	M00001342B:E06
312		RTA00000187AF.h.23.1	M00001679A:F06
313	6298	RTA00000187AR.i.10.2	M00001679B:F01

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
314	14367	RTA00000187AF.e.8.1	M00001670C:H02
315		RTA00000193AF.c.22.1	M00004249D:G12
316	16921	RTA00000183AF.k.6.1	M00001534A:C04
317	1577	RTA00000184AF.i.23.1	M00001556A:F11
318	8773	RTA00000187AF.f.24.1	M00001675A:C09
319		RTA00000194AF.a.11.1	M00004461A:B09
320	39886	RTA00000178AF.j.24.1	M00001380D:B09
321	13532	RTA00000181AF.c.4.1	M00001445A:F05
322		RTA00000193AF.d.2.1	M00004251C:G07
323	5257	RTA00000192AF.f.3.1	M00004146C:C11
324	9061	RTA00000191AR.e.11.2	M00004031A:A12
325	19267	RTA00000186AF.l.12.1	M00001645A:C12
326	20212	RTA00000134A.l.22.1	M00001535A:C06
327	16653	RTA00000181AF.k.5.3	M00001453C:F06
328	16985	RTA00000177AF.h.10.1	M00001348B:G06
329	12977	RTA00000189AR.j.19.1	M00003875B:F04
330	9061	RTA00000191AR.e.11.3	M00004031A:A12
331		RTA00000194AR.a.10.2	M00004461A:B08
332	6468	RTA00000187AF.d.15.1	M00001669B:F02
333	16392	RTA00000192AF.l.1.1	M00004183C:D07
334	14627	RTA00000187AF.g.23.1	M00001677C:E10
335	6583	RTA00000179AF.d.13.1	M00001394A:F01
336	6806	RTA00000177AF.g.13.3	M00001346D:E03
337	9635	RTA00000137A.e.23.4	M00001557A:F01
338	689	RTA00000181AR.l.22.1	M00001454D:G03
339	4119	RTA00000183AF.k.16.1	M00001534C:A01
340	8952	RTA00000183AF.h.15.1	M00001518C:B11
341	2379	RTA00000192AF.p.8.1	M00004212B:C07
342	39486	RTA00000128A.m.22.2	M00001561A:C05
343	21877	RTA00000189AF.b.21.1	M00003833A:E05
344	6874	RTA00000192AF.a.14.1	M00004111D:A08
345	6874	RTA00000189AF.e.9.1	M00003846B:D06
346	37285	RTA00000191AF.f.11.1	M00004035C:A07
347		RTA00000193AF.j.20.1	M00004327B:H04
348	7674	RTA00000118A.g.9.1	M00001416A:H01
349	2797	RTA00000180AF.i.19.1	M00001429A:H04
350		RTA00000184AF.g.22.1	M00001552D:A01

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
351	7802	RTA00000185AF.n.5.1	M00001608A:B03
352	16921	RTA00000193AF.h.15.1	M00004295D:F12
353	11494	RTA00000192AF.j.6.1	M00004172C:D08
354	17062	RTA00000177AF.b.8.4	M00001340B:A06
355	16245	RTA00000177AF.k.9.3	M00001352A:E02
356	83103	RTA00000119A.e.24.2	M00001454A:A09
357	4309	RTA00000186AF.e.22.1	M00001624C:F01
358	13072	RTA00000181AR.m.5.2	M00001455B:E12
359	4059	RTA00000177AF.n.18.3	M00001357D:D11
360	5178	RTA00000178AF.n.10.1	M00001386C:B12
361	1120	RTA00000118A.p.15.3	M00001452A:D08
362	6420	RTA00000183AF.d.11.1	M00001504D:G06
363	13913	RTA00000186AF.e.6.1	M00001623D:F10
364		RTA00000192AF.c.2.1	M00004121B:G01
365	3956	RTA00000183AF.g.3.1	M00001512D:G09
366	14364	RTA00000183AF.g.12.1	M00001513C:E08
367	6880	RTA00000191AF.m.20.1	M00004087D:A01
368	84182	RTA00000180AF.h.19.1	M00001428A:H10
369	2790	RTA00000177AF.e.2.1	M00001343C:F10
370	4561	RTA00000184AF.i.21.1	M00001555D:G10
371	8847	RTA00000180AF.b.16.1	M00001416B:H11
372	56020	RTA00000193AF.g.2.1	M00004285B:E08
373	1531	RTA00000119A.o.3.1	M00001461A:D06
374	6420	RTA00000177AF.f.10.3	M00001345A:E01
375		RTA00000188AF.b.12.1	M00003754C:E09
376		RTA00000180AF.k.24.1	M00001432C:F06
377		RTA00000184AF.a.8.1	M00001544A:E06
378	2696	RTA00000134A.m.13.1	M00001536A:B07
379	260	RTA00000185AR.i.12.2	M00001594B:H04
380	11350	RTA00000189AF.a.24.2	M00003826B:A06
381	2428	RTA00000123A.l.21.1	M00001533A:C11
382	4313	RTA00000122A.n.3.1	M00001517A:B07
383		RTA00000184AF.p.3.1	M00001566B:D11
384	697	RTA00000188AF.d.6.1	M00003759B:B09
385	5619	RTA00000188AF.l.9.1	M00003796C:D05
386	4568	RTA00000122A.d.15.3	M00001513A:B06
387		RTA00000177AF.i.6.2	M00001350A:B08

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
388	5622	RTA00000178AF.a.11.1	M00001362B:D10
389	7514	RTA00000184AF.k.21.1	M00001558B:H11
390	5619	RTA00000189AF.f.17.1	M00003853A:D04
391	7570	RTA00000187AF.g.24.1	M00001677D:A07
392	23358	RTA00000190AF.o.21.1	M00003974D:H02
393	23210	RTA00000190AF.o.20.1	M00003974D:E07
394	5192	RTA00000184AF.k.2.1	M00001557B:H10
395	13538	RTA00000180AF.a.24.1	M00001415A:H06
396		RTA00000189AF.h.17.1	M00003867A:D10
397		RTA00000192AF.o.11.1	M00004205D:F06
398		RTA00000184AF.l.11.1	M00001559B:F01
399	4718	RTA00000189AF.g.5.1	M00003857A:H03
400	14929	RTA00000177AF.m.1.2	M00001353D:D10
401	4908	RTA00000192AF.j.2.1	M00004171D:B03
402		RTA00000178AF.k.16.1	M00001381D:E06
403		RTA00000194AF.c.24.1	M00004692A:H08
404	17732	RTA00000178AR.i.2.2	M00001376B:G06
405	17062	80.A1.sp6:130208.Seq	M00001340B:A06
406	11589	80.B1.sp6:130220.Seq	M00001340D:F10
407	4443	80.C1.sp6:130232.Seq	M00001341A:E12
408	39805	80.D1.sp6:130244.Seq	M00001342B:E06
409	2790	80.E1.sp6:130256.Seq	M00001343C:F10
410	23255	80.F1.sp6:130268.Seq	M00001343D:H07
411	6420	80.G1.sp6:130280.Seq	M00001345A:E01
412	5007	80.H1.sp6:130292.Seq	M00001346A:F09
413	13576	80.D2.sp6:130245.Seq	M00001347A:B10
414	16927	80.E2.sp6:130257.Seq	M00001348B:B04
415	16985	80.F2.sp6:130269.Seq	M00001348B:G06
416	3584	80.G2.sp6:130281.Seq	M00001349B:B08
417		80.H2.sp6:130293.Seq	M00001350A:B08
418	7187	80.A3.sp6:130210.Seq	M00001350A:H01
419	16245	80.D3.sp6:130246.Seq	M00001352A:E02
420	8078	80.E3.sp6:130258.Seq	M00001353A:G12
421	14929	80.F3.sp6:130270.Seq	M00001353D:D10
422	14391	80.G3.sp6:130282.Seq	M00001355B:G10
423	4141	80.B4.sp6:130223.Seq	M00001361A:A05
424	2379	80.C4.sp6:130235.Seq	M00001361D:F08

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
425	5622	80.D4.sp6:130247.Seq	M00001362B:D10
426	945	80.E4.sp6:130259.Seq	M00001362C:H11
427	40132	80.F4.sp6:130271.Seq	M00001365C:C10
428		80.G4.sp6:130283.Seq	M00001368D:E03
429	6867	80.H4.sp6:130295.Seq	M00001370A:C09
430	7172	80.A5.sp6:130212.Seq	M00001371C:E09
431	17732	80.B5.sp6:130224.Seq	M00001376B:G06
432	39833	80.C5.sp6:130236.Seq	M00001378B:B02
433	1334	80.D5.sp6:130248.Seq	M00001379A:A05
434	39886	80.E5.sp6:130260.Seq	M00001380D:B09
435		80.F5.sp6:130272.Seq	M00001381D:E06
436	22979	80.G5.sp6:130284.Seq	M00001382C:A02
437	39648	80.H5.sp6:130296.Seq	M00001383A:C03
438		80.B6.sp6:130225.Seq	M00001384B:A11
439	5178	80.C6.sp6:130237.Seq	M00001386C:B12
440	2464	80.D6.sp6:130249.Seq	M00001387A:C05
441	7587	80.E6.sp6:130261.Seq	M00001387B:G03
442	5832	80.F6.sp6:130273.Seq	M00001388D:G05
443	16269	80.G6.sp6:130285.Seq	M00001389A:C08
444	6583	80.H6.sp6:130297.Seq	M00001394A:F01
445	4009	80.A7.sp6:130214.Seq	M00001396A:C03
446		80.B7.sp6:130226.Seq	M00001400B:H06
447	39563	80.C7.sp6:130238.Seq	M00001402A:E08
448	5556	80.D7.sp6:130250.Seq	M00001407B:D11
449	9577	80.E7.sp6:130262.Seq	M00001409C:D12
450	7005	80.F7.sp6:130274.Seq	M00001410A:D07
451	8551	80.G7.sp6:130286.Seq	M00001412B:B10
452		80.H7.sp6:130298.Seq	M00001414A:B01
453		80.A8.sp6:130215.Seq	M00001414C:A07
454	13538	80.B8.sp6:130227.Seq	M00001415A:H06
455	8847	80.C8.sp6:130239.Seq	M00001416B:H11
456	36393	80.D8.sp6:130251.Seq	M00001417A:E02
457	9952	80.E8.sp6:130263.Seq	M00001418B:F03
458	9577	80.G8.sp6:130287.Seq	M00001421C:F01
459	15066	80.H8.sp6:130299.Seq	M00001423B:E07
460	10470	80.A9.sp6:130216.Seq	M00001424B:G09
461	22195	80.B9.sp6:130228.Seq	M00001425B:H08

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
462		80.C9.sp6:130240.Seq	M00001426B:D12
463	4261	80.D9.sp6:130252.Seq	M00001426D:C08
464	84182	80.E9.sp6:130264.Seq	M00001428A:H10
465	40392	80.H9.sp6:130300.Seq	M00001429D:D07
466	16731	80.C10.sp6:130241.Seq	M00001442C:D07
467		80.D10.sp6:130253.Seq	M00001443B:F01
468	13532	80.E10.sp6:130265.Seq	M00001445A:F05
469	8	80.H10.sp6:130301.Seq	M00001448D:C09
470	36313	80.A11.sp6:130218.Seq	M00001448D:H01
471	5857	80.B11.sp6:130230.Seq	M00001449A:A12
472	41633	80.C11.sp6:130242.Seq	M00001449A:B12
473	36535	80.D11.sp6:130254.Seq	M00001449A:G10
474	86110	80.E11.sp6:130266.Seq	M00001449C:D06
475	32663	80.F11.sp6:130278.Seq	M00001450A:A11
476	27250	80.G11.sp6:130290.Seq	M00001450A:D08
477	16970	80.H11.sp6:130302.Seq	M00001452C:B06
478	16130	80.A12.sp6:130219.Seq	M00001453A:E11
479	16653	80.B12.sp6:130231.Seq	M00001453C:F06
480	7005	80.C12.sp6:130243.Seq	M00001454B:C12
481	13072	80.F12.sp6:130279.Seq	M00001455B:E12
482	9283	80.G12.sp6:130291.Seq	M00001455D:F09
483	23255	100.C1.sp6:131446.Seq	M00001343D:H07
484	13576	100.E1.sp6:131470.Seq	M00001347A:B10
485	7187	100.C2.sp6:131447.Seq	M00001350A:H01
486	14391	100.E3.sp6:131472.Seq	M00001355B:G10
487	945	100.E4.sp6:131473.Seq	M00001362C:H11
488	7172	100.A5.sp6:131426.Seq	M00001371C:E09
489	39648	100.A6.sp6:131427.Seq	M00001383A:C03
490	84182	100.G9.sp6:131502.Seq	M00001428A:H10
491	8	100.B11.sp6:131444.Seq	M00001448D:C09
492	36535	100.D11.sp6:131468.Seq	M00001449A:G10
493	82498	100.F11.sp6:131492.Seq	M00001450A:B12
494	16970	100.C12.sp6:131457.Seq	M00001452C:B06
495	16130	100.D12.sp6:131469.Seq	M00001453A:E11
496	7005	121.D1.sp6:131917.Seq	M00001454B:C12
497		121.G6.sp6:131958.Seq	M00001506D:A09
498	18957	121.F7.sp6:131947.Seq	M00001528A:F09

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
499	40044	122.E1.sp6:132121.Seq	M00001621C:C08
500	5214	122.C2.sp6:132098.Seq	M00001630B:H09
501	6660	122.B5.sp6:132089.Seq	M00001679A:A06
502	13183	123.D5.sp6:132305.Seq	M00004114C:F11
503	6455	123.E7.sp6:132319.Seq	M00004157C:A09
504	5319	123.F7.sp6:132331.Seq	M00004169C:C12
505	11443	123.A8.sp6:132272.Seq	M00004185C:C03
506		123.C8.sp6:132296.Seq	M00004191D:B11
507	8210	123.E8.sp6:132320.Seq	M00004197D:H01
508	9457	123.D11.sp6:132311.Seq	M00004307C:A06
509	6420	172.E1.sp6:133925.Seq	M00001345A:E01
510	16245	172.D2.sp6:133914.Seq	M00001352A:E02
511	8078	172.C3.sp6:133903.Seq	M00001353A:G12
512	14929	172.D3.sp6:133915.Seq	M00001353D:D10
513	14391	172.H3.sp6:133963.Seq	M00001355B:G10
514	6583	172.B8.sp6:133896.Seq	M00001394A:F01
515	4009	172.D8.sp6:133920.Seq	M00001396A:C03
516		172.B9.sp6:133897.Seq	M00001400B:H06
517		176.A3.sp6:134514.Seq	M00001632D:H07
518	19267	176.G3.sp6:134586.Seq	M00001645A:C12
519	78091	176.G5.sp6:134588.Seq	M00001679C:F01
520	17055	176.D6.sp6:134553.Seq	M00001682C:B12
521	6539	176.D9.sp6:134556.Seq	M00003844C:B11
522		177.H4.sp6:134791.Seq	M00004121B:G01
523	5257	177.F5.sp6:134768.Seq	M00004146C:C11
524	11494	177.E6.sp6:134757.Seq	M00004172C:D08
525		177.G7.sp6:134782.Seq	M00004205D:F06
526	11451	177.D8.sp6:134747.Seq	M00004214C:H05
527	9283	173.D2.SP6:134106.Seq	M00001455D:F09
528	16283	173.F3.SP6:134131.Seq	M00001467A:D08
529	10539	173.B5.SP6:134085.Seq	M00001499B:A11
530	6420	173.F5.SP6:134133.Seq	M00001504D:G06
531	3956	173.H5.SP6:134157.Seq	M00001512D:G09
532		173.G7.SP6:134147.Seq	M00001544A:E06
533	1577	173.C9.SP6:134101.Seq	M00001556A:F11
534	9635	173.D9.SP6:134113.Seq	M00001557A:F01
535	5192	173.E9.SP6:134125.Seq	M00001557B:H10

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
536	6539	173.A12.SP6:134080.Seq	M00001579D:C03
537	945	180.C2.sp6:135940.Seq	M00001362C:H11
538	7005	180.H5.sp6:136003.Seq	M00001410A:D07
539	39304	180.G9.sp6:135995.Seq	M00001450A:A02
540	27250	180.B10.sp6:135936.Seq	M00001450A:D08
541	35555	184.A5.sp6:135530.Seq	M00001528A:C04
542	19255	184.B10.sp6:135547.Seq	M00001545A:C03
543	6268	184.C12.sp6:135561.Seq	M00001551A:B10
544	3277	217.E1.sp6:139406.Seq	M00001624A:B06
545	39171	217.A12.sp6:139369.Seq	M00001644C:B07
546	11460	219.F2.sp6:139035.Seq	M00001676B:F05
547	10539	219.F6.sp6:139039.Seq	M00001680D:F08
548	11476	219.H8.sp6:139065.Seq	M00003747D:C05
549	4016	79.A1.sp6:130016.Seq	M00001395A:C03
550	7674	79.C1.sp6:130040.Seq	M00001416A:H01
551	3681	79.E1.sp6:130064.Seq	M00001449A:D12
552	39304	79.F1.sp6:130076.Seq	M00001450A:A02
553	82498	79.G1.sp6:130088.Seq	M00001450A:B12
554	84328	79.A2.sp6:130017.Seq	M00001452A:B04
555	86859	79.B2.sp6:130029.Seq	M00001452A:B12
556	1120	79.C2.sp6:130041.Seq	M00001452A:D08
557	85064	79.D2.sp6:130053.Seq	M00001452A:F05
558	83103	79.G2.sp6:130089.Seq	M00001454A:A09
559	10145	79.F3.sp6:130078.Seq	M00001465A:B11
560	16283	79.H3.sp6:130102.Seq	M00001467A:D08
561	4568	79.D4.sp6:130055.Seq	M00001513A:B06
562	4313	79.F4.sp6:130079.Seq	M00001517A:B07
563	2428	79.A5.sp6:130020.Seq	M00001533A:C11
564	39423	79.C5.sp6:130044.Seq	M00001535A:F10
565	39174	79.E5.sp6:130068.Seq	M00001541A:H03
566	22113	79.F5.sp6:130080.Seq	M00001542A:A09
567	19829	79.H5.sp6:130104.Seq	M00001544A:G02
568	13864	79.B6.sp6:130033.Seq	M00001545A:D08
569	1058	79.F6.sp6:130081.Seq	M00001548A:H09
570	4015	79.G6.sp6:130093.Seq	M00001549A:B02
571	39180	79.A7.sp6:130022.Seq	M00001551A:F05
572	307	79.C7.sp6:130046.Seq	M00001552A:B12

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
573	39458	79.D7.sp6:130058.Seq	M00001552A:D11
574	39490	79.G7.sp6:130094.Seq	M00001557A:F03
575	39486	79.B8.sp6:130035.Seq	M00001561A:C05
576	39380	79.E8.sp6:130071.Seq	M00001587A:B11
577	1399	79.G8.sp6:130095.Seq	M00001604A:B10
578	39391	79.A9.sp6:130024.Seq	M00001604A:F05
579	6268	79.G9.sp6:130096.Seq	M00001551A:B10
580		377.F4.sp6:141957.Seq	M00004692A:H08
581	2448	89.A1.sp6:130667.Seq	M00001460A:F06
582	1531	89.C1.sp6:130691.Seq	M00001461A:D06
583	19	89.D1.sp6:130703.Seq	M00001463C:B11
584	38759	89.F1.sp6:130727.Seq	M00001467A:B07
585	39508	89.G1.sp6:130739.Seq	M00001467A:D04
586	16283	89.H1.sp6:130751.Seq	M00001467A:D08
587	39442	89.A2.sp6:130668.Seq	M00001467A:E10
588	7589	89.B2.sp6:130680.Seq	M00001468A:F05
589		89.C2.sp6:130692.Seq	M00001469A:A01
590	12081	89.D2.sp6:130704.Seq	M00001469A:C10
591	19105	89.E2.sp6:130716.Seq	M00001469A:H12
592	1037	89.F2.sp6:130728.Seq	M00001470A:B10
593	39425	89.G2.sp6:130740.Seq	M00001470A:C04
594	39478	89.H2.sp6:130752.Seq	M00001471A:B01
595		89.B3.sp6:130681.Seq	M00001487B:H06
596		89.C3.sp6:130693.Seq	M00001488B:F12
597	18699	89.D3.sp6:130705.Seq	M00001490B:C04
598	7206	89.E3.sp6:130717.Seq	M00001494D:F06
599	2623	89.F3.sp6:130729.Seq	M00001497A:G02
600	10539	89.G3.sp6:130741.Seq	M00001499B:A11
601	5336	89.H3.sp6:130753.Seq	M00001500A:C05
602	2623	89.A4.sp6:130670.Seq	M00001500A:E11
603	9443	89.B4.sp6:130682.Seq	M00001500C:E04
604	9685	89.C4.sp6:130694.Seq	M00001501D:C02
605		89.D4.sp6:130706.Seq	M00001504A:E01
606	10185	89.E4.sp6:130718.Seq	M00001504C:A07
607	6974	89.F4.sp6:130730.Seq	M00001504C:H06
608	6420	89.G4.sp6:130742.Seq	M00001504D:G06
609		89.H4.sp6:130754.Seq	M00001505C:C05

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
610		89.A5.sp6:130671.Seq	M00001506D:A09
611	39168	89.B5.sp6:130683.Seq	M00001507A:H05
612	39412	89.C5.sp6:130695.Seq	M00001511A:H06
613	39186	89.D5.sp6:130707.Seq	M00001512A:A09
614	3956	89.E5.sp6:130719.Seq	M00001512D:G09
615		89.F5.sp6:130731.Seq	M00001513B:G03
616	14364	89.G5.sp6:130743.Seq	M00001513C:E08
617	40044	89.H5.sp6:130755.Seq	M00001514C:D11
618	8952	89.A6.sp6:130672.Seq	M00001518C:B11
619	35555	89.B6.sp6:130684.Seq	M00001528A:C04
620	18957	89.C6.sp6:130696.Seq	M00001528A:F09
621	8358	89.D6.sp6:130708.Seq	M00001528B:H04
622	38085	89.E6.sp6:130720.Seq	M00001531A:D01
623		89.F6.sp6:130732.Seq	M00001531A:H11
624	3990	89.G6.sp6:130744.Seq	M00001532B:A06
625	16921	89.H6.sp6:130756.Seq	M00001534A:C04
626	5321	89.B7.sp6:130685.Seq	M00001534A:F09
627	4119	89.C7.sp6:130697.Seq	M00001534C:A01
628	20212	89.E7.sp6:130721.Seq	M00001535A:C06
629	2696	89.F7.sp6:130733.Seq	M00001536A:B07
630	39392	89.G7.sp6:130745.Seq	M00001536A:C08
631	39420	89.H7.sp6:130757.Seq	M00001537A:F12
632	3389	89.A8.sp6:130674.Seq	M00001537B:G07
633	8286	89.B8.sp6:130686.Seq	M00001540A:D06
634	3765	89.C8.sp6:130698.Seq	M00001541A:D02
635	39453	89.E8.sp6:130722.Seq	M00001542A:E06
636		89.F8.sp6:130734.Seq	M00001542B:B01
637		89.H8.sp6:130758.Seq	M00001544A:E06
638	6974	89.A9.sp6:130675.Seq	M00001544B:B07
639		89.B9.sp6:130687.Seq	M00001545A:B02
640	19255	89.C9.sp6:130699.Seq	M00001545A:C03
641	1267	89.D9.sp6:130711.Seq	M00001546A:G11
642	5892	89.E9.sp6:130723.Seq	M00001548A:E10
643	4193	89.G9.sp6:130747.Seq	M00001549B:F06
644	16347	89.H9.sp6:130759.Seq	M00001549C:E06
645	7239	89.A10.sp6:130676.Seq	M00001550A:A03
646	5175	89.B10.sp6:130688.Seq	M00001550A:G01

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
647	22390	89.C10.sp6:130700.Seq	M00001551A:G06
648	3266	89.D10.sp6:130712.Seq	M00001551C:G09
649	5708	89.E10.sp6:130724.Seq	M00001552B:D04
650		89.F10.sp6:130736.Seq	M00001552D:A01
651	8298	89.G10.sp6:130748.Seq	M00001553A:H06
652	4573	89.H10.sp6:130760.Seq	M00001553B:F12
653	22814	89.A11.sp6:130677.Seq	M00001553D:D10
654	39539	89.B11.sp6:130689.Seq	M00001555A:B02
655	39195	89.C11.sp6:130701.Seq	M00001555A:C01
656	4561	89.D11.sp6:130713.Seq	M00001555D:G10
657	9244	89.E11.sp6:130725.Seq	M00001556A:C09
658	1577	89.F11.sp6:130737.Seq	M00001556A:F11
659	4386	89.H11.sp6:130761.Seq	M00001556B:C08
660	11294	89.A12.sp6:130678.Seq	M00001556B:G02
661	5192	89.D12.sp6:130714.Seq	M00001557B:H10
662	8761	89.E12.sp6:130726.Seq	M00001557D:D09
663		89.F12.sp6:130738.Seq	M00001558A:H05
664	7514	89.G12.sp6:130750.Seq	M00001558B:H11
665		89.H12.sp6:130762.Seq	M00001559B:F01
666	6558	90.A1.sp6:130859.Seq	M00001560D:F10
667	102	90.B1.sp6:130871.Seq	M00001563B:F06
668		90.D1.sp6:130895.Seq	M00001566B:D11
669	5749	90.E1.sp6:130907.Seq	M00001571C:H06
670	6539	90.G1.sp6:130931.Seq	M00001579D:C03
671	6293	90.A2.sp6:130860.Seq	M00001583D:A10
672		90.C2.sp6:130884.Seq	M00001590B:F03
673	260	90.D2.sp6:130896.Seq	M00001594B:H04
674	4837	90.E2.sp6:130908.Seq	M00001597C:H02
675	10470	90.F2.sp6:130920.Seq	M00001597D:C05
676	16999	90.G2.sp6:130932.Seq	M00001598A:G03
677	22794	90.H2.sp6:130944.Seq	M00001601A:D08
678	11465	90.A3.sp6:130861.Seq	M00001607A:E11
679	7802	90.B3.sp6:130873.Seq	M00001608A:B03
680	22155	90.C3.sp6:130885.Seq	M00001608B:E03
681		90.D3.sp6:130897.Seq	M00001608D:A11
682	13157	90.E3.sp6:130909.Seq	M00001614C:F10
683	17004	90.F3.sp6:130921.Seq	M00001617C:E02

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
684	40314	90.G3.sp6:130933.Seq	M00001619C:F12
685	40044	90.H3.sp6:130945.Seq	M00001621C:C08
686	13913	90.A4.sp6:130862.Seq	M00001623D:F10
687	3277	90.B4.sp6:130874.Seq	M00001624A:B06
688	4309	90.C4.sp6:130886.Seq	M00001624C:F01
689	5214	90.D4.sp6:130898.Seq	M00001630B:H09
690		90.E4.sp6:130910.Seq	M00001632D:H07
691	39171	90.F4.sp6:130922.Seq	M00001644C:B07
692	19267	90.G4.sp6:130934.Seq	M00001645A:C12
693	4665	90.H4.sp6:130946.Seq	M00001648C:A01
694		90.A5.sp6:130863.Seq	M00001651A:H01
695	23201	90.B5.sp6:130875.Seq	M00001657D:C03
696	76760	90.C5.sp6:130887.Seq	M00001657D:F08
697	23218	90.D5.sp6:130899.Seq	M00001662C:A09
698	35702	90.E5.sp6:130911.Seq	M00001663A:E04
699	6468	90.F5.sp6:130923.Seq	M00001669B:F02
700	14367	90.G5.sp6:130935.Seq	M00001670C:H02
701	7015	90.H5.sp6:130947.Seq	M00001673C:H02
702	8773	90.A6.sp6:130864.Seq	M00001675A:C09
703	11460	90.B6.sp6:130876.Seq	M00001676B:F05
704	7570	90.D6.sp6:130900.Seq	M00001677D:A07
705	4416	90.E6.sp6:130912.Seq	M00001678D:F12
706	6660	90.F6.sp6:130924.Seq	M00001679A:A06
707		90.H6.sp6:130948.Seq	M00001679A:F06
708	26875	90.A7.sp6:130865.Seq	M00001679A:F10
709	6298	90.B7.sp6:130877.Seq	M00001679B:F01
710	78091	90.C7.sp6:130889.Seq	M00001679C:F01
711	10751	90.D7.sp6:130901.Seq	M00001679D:D03
712	10539	90.F7.sp6:130925.Seq	M00001680D:F08
713	17055	90.G7.sp6:130937.Seq	M00001682C:B12
714	5382	90.A8.sp6:130866.Seq	M00001688C:F09
715	4393	90.B8.sp6:130878.Seq	M00001693C:G01
716	67252	90.C8.sp6:130890.Seq	M00001716D:H05
717	40108	90.D8.sp6:130902.Seq	M00003741D:C09
718	11476	90.E8.sp6:130914.Seq	M00003747D:C05
719		90.F8.sp6:130926.Seq	M00003754C:E09
720	697	90.G8.sp6:130938.Seq	M00003759B:B09

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
721		90.H8.sp6:130950.Seq	M00003761D:A09
722	17076	90.A9.sp6:130867.Seq	M00003762C:B08
723	3108	90.B9.sp6:130879.Seq	M00003763A:F06
724	67907	90.C9.sp6:130891.Seq	M00003774C:A03
725		90.D9.sp6:130903.Seq	M00003784D:D12
726	11350	90.F9.sp6:130927.Seq	M00003826B:A06
727	7899	90.H9.sp6:130951.Seq	M00003837D:A01
728	7798	90.A10.sp6:130868.Seq	M00003839A:D08
729	6539	90.B10.sp6:130880.Seq	M00003844C:B11
730	6874	90.C10.sp6:130892.Seq	M00003846B:D06
731		90.D10.sp6:130904.Seq	M00003851B:D08
732	13595	90.E10.sp6:130916.Seq	M00003851B:D10
733	5619	90.F10.sp6:130928.Seq	M00003853A:D04
734	10515	90.G10.sp6:130940.Seq	M00003853A:F12
735	4622	90.H10.sp6:130952.Seq	M00003856B:C02
736	3389	90.A11.sp6:130869.Seq	M00003857A:G10
737	4718	90.B11.sp6:130881.Seq	M00003857A:H03
738		90.C11.sp6:130893.Seq	M00003867A:D10
739	12977	90.F11.sp6:130929.Seq	M00003875B:F04
740	8479	90.G11.sp6:130941.Seq	M00003875C:G07
741		90.H11.sp6:130953.Seq	M00003875D:D11
742	7798	90.A12.sp6:130870.Seq	M00003876D:E12
743	5345	90.B12.sp6:130882.Seq	M00003879B:C11
744	31587	90.C12.sp6:130894.Seq	M00003879B:D10
745	14507	90.D12.sp6:130906.Seq	M00003879D:A02
746	13576	90.F12.sp6:130930.Seq	M00003885C:A02
747		90.G12.sp6:130942.Seq	M00003891C:H09
748	9285	90.H12.sp6:130954.Seq	M00003906C:E10
749	39809	99.A1.sp6:131230.Seq	M00003907D:A09
750	16317	99.B1.sp6:131242.Seq	M00003907D:H04
751	8672	99.C1.sp6:131254.Seq	M00003909D:C03
752	12532	99.D1.sp6:131266.Seq	M00003912B:D01
753	3900	99.E1.sp6:131278.Seq	M00003914C:F05
754	23255	99.F1.sp6:131290.Seq	M00003922A:E06
755	24488	99.C2.sp6:131255.Seq	M00003968B:F06
756	40122	99.D2.sp6:131267.Seq	M00003970C:B09
757	23210	99.E2.sp6:131279.Seq	M00003974D:E07

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
758	23358	99.F2.sp6:131291.Seq	M00003974D:H02
759	3430	99.A3.sp6:131232.Seq	M00003981A:E10
760	2433	99.B3.sp6:131244.Seq	M00003982C:C02
761	9105	99.C3.sp6:131256.Seq	M00003983A:A05
762	6124	99.D3.sp6:131268.Seq	M00004028D:A06
763	40073	99.E3.sp6:131280.Seq	M00004028D:C05
764	37285	99.H3.sp6:131316.Seq	M00004035C:A07
765	17036	99.A4.sp6:131233.Seq	M00004035D:B06
766	3706	99.C4.sp6:131257.Seq	M00004068B:A01
767		99.D4.sp6:131269.Seq	M00004072A:C03
768	15069	99.F4.sp6:131293.Seq	M00004081C:D10
769	9285	99.H4.sp6:131317.Seq	M00004086D:G06
770	6880	99.A5.sp6:131234.Seq	M00004087D:A01
771	5325	99.C5.sp6:131258.Seq	M00004093D:B12
772	7221	99.D5.sp6:131270.Seq	M00004105C:A04
773	4937	99.E5.sp6:131282.Seq	M00004108A:E06
774	6874	99.F5.sp6:131294.Seq	M00004111D:A08
775	13183	99.G5.sp6:131306.Seq	M00004114C:F11
776		99.H5.sp6:131318.Seq	M00004121B:G01
777	13272	99.A6.sp6:131235.Seq	M00004138B:H02
778	5257	99.B6.sp6:131247.Seq	M00004146C:C11
779	6455	99.D6.sp6:131271.Seq	M00004157C:A09
780	5319	99.E6.sp6:131283.Seq	M00004169C:C12
781	4908	99.F6.sp6:131295.Seq	M00004171D:B03
782	11494	99.G6.sp6:131307.Seq	M00004172C:D08
783	11443	99.A7.sp6:131236.Seq	M00004185C:C03
784		99.B7.sp6:131248.Seq	M00004191D:B11
785	8210	99.C7.sp6:131260.Seq	M00004197D:H01
786	14311	99.D7.sp6:131272.Seq	M00004203B:C12
787		99.E7.sp6:131284.Seq	M00004205D:F06
788	12971	99.B8.sp6:131249.Seq	M00004223D:E04
789	6455	99.C8.sp6:131261.Seq	M00004229B:F08
790	7212	99.D8.sp6:131273.Seq	M00004230B:C07
791	4905	99.H8.sp6:131321.Seq	M00004269D:D06
792	16914	99.A9.sp6:131238.Seq	M00004275C:C11
793	16921	99.D9.sp6:131274.Seq	M00004295D:F12
794	13046	99.E9.sp6:131286.Seq	M00004296C:H07

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
795	9457	99.F9.sp6:131298.Seq	M00004307C:A06
796	26295	99.G9.sp6:131310.Seq	M00004312A:G03
797	21847	99.H9.sp6:131322.Seq	M00004318C:D10
798		99.H10.sp6:131323.Seq	M00004505D:F08
799		99.B11.sp6:131252.Seq	M00004692A:H08
800		99.D11.sp6:131276.Seq	M00005180C:G03
801	39304	RTA00000118A.j.21.1.Seq_THC151859	
802	2428	RTA00000123A.l.21.1.Seq_THC205063	
803	1058	RTA00000126A.e.20.3.Seq_THC217534	
804	5097	RTA00000134A.k.1.1.Seq_THC215869	
805	20212	RTA00000134A.l.22.1.Seq_THC128232	
806	23255	RTA00000177AF.e.14.3.Seq_THC228776	
807	2790	RTA00000177AF.e.2.1.Seq_THC229461	
808	6420	RTA00000177AF.f.10.3.Seq_THC226443	
809	4059	RTA00000177AF.n.18.3.Seq_THC123051	
810		RTA00000179AF.j.13.1.Seq_THC105720	
811	9952	RTA00000180AF.c.20.1.Seq_THC162284	
812	13238	RTA00000181AF.m.4.1.Seq_THC140691	
813	9685	RTA00000183AF.c.11.1.Seq_THC109544	
814		RTA00000183AF.c.24.1.Seq_THC125912	
815	6420	RTA00000183AF.d.11.1.Seq_THC226443	
816	6974	RTA00000183AF.d.9.1.Seq_THC223129	
817	40044	RTA00000183AF.g.22.1.Seq_THC232899	
818		RTA00000183AF.g.9.1.Seq_THC198280	
819	5892	RTA00000184AF.d.11.1.Seq_THC161896	
820	40044	RTA00000186AF.d.1.1.Seq_THC232899	
821		RTA00000186AF.h.14.1.Seq_THC112525	
822	19267	RTA00000186AF.l.12.1.Seq_THC178183	
823	8773	RTA00000187AF.f.24.1.Seq_THC220002	
824	7570	RTA00000187AF.g.24.1.Seq_THC168636	
825	11476	RTA00000187AF.p.19.1.Seq_THC108482	
826		RTA00000188AF.d.11.1.Seq_THC212094	
827	17076	RTA00000188AF.d.21.1.Seq_THC208760	
828	697	RTA00000188AF.d.6.1.Seq_THC178884	
829	67907	RTA00000188AF.g.11.1.Seq_THC123222	
830	5619	RTA00000188AF.l.9.1.Seq_THC167845	
831	4718	RTA00000189AF.g.5.1.Seq_THC196102	

SEQ ID NO:	Cluster ID	Sequence Name	Clone Name
832	39809	RTA00000190AF.e.3.1.Seq_THC150217	
833	23255	RTA00000190AF.j.4.1.Seq_THC228776	
834	40122	RTA00000190AF.n.23.1.Seq_THC109227	
835	23210	RTA00000190AF.o.20.1.Seq_THC207240	
836	23358	RTA00000190AF.o.21.1.Seq_THC207240	
837	5693	RTA00000190AF.p.17.2.Seq_THC173318	
838	2433	RTA00000191AF.a.15.2.Seq_THC79498	
839	5257	RTA00000192AF.f.3.1.Seq_THC213833	
840	16392	RTA00000192AF.l.1.1.Seq_THC202071	
841		RTA00000193AF.c.21.1.Seq_THC222602	
842	26295	RTA00000193AF.i.24.2.Seq_THC197345	
843		RTA00000193AF.m.5.1.Seq_THC173318	
844		RTA00000193AF.n.15.1.Seq_THC215687	

Example 2: Results of Public Database Search to Identify Function of Gene Products

SEQ ID NOS:1-404, as well as the validation sequences SEQ ID NOS:405-800, were translated in all three reading frames to determine the best alignment with the individual sequences. These amino acid sequences and nucleotide sequences are referred, generally, as query sequences, which are aligned with the individual sequences. Query and individual sequences were aligned using the BLAST programs, available over the world wide web sit of the NCBI.. Again the sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above in Example 1.

Table 2 (inserted before the claims) shows the results of the alignments. Table 2 refers to each sequence by its SEQ ID NO:, the accession numbers and descriptions of nearest neighbors from the Genbank and Non-Redundant Protein searches, and the p values of the search results. Table 1 identifies each SEQ ID NO: by SEQ name, clone ID, and cluster. As discussed above, a single cluster includes polynucleotides representing the same gene or gene family, and generally represents sequences encoding the same gene product.

For each of SEQ ID NOS:1-800, the best alignment to a protein or DNA sequence is included in Table 2. The activity of the polypeptide encoded by SEQ ID NOS:1-800 is the same or similar to the nearest neighbor reported in Table 2. The accession number of the nearest neighbor is reported, providing a reference to the activities exhibited by the nearest neighbor. The search program and database used for the alignment also are indicated as well as a calculation of the p value.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence of SEQ ID NOS:1-800. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of SEQ ID NOS:1-800.

SEQ ID NOS:1-800 and the translations thereof may be human homologs of known genes of other species or novel allelic variants of known human genes. In such cases, these new human sequences are suitable as diagnostics or therapeutics. As diagnostics, the human sequences SEQ ID NOS:1-800 exhibit greater specificity in detecting and differentiating human cell lines and types than homologs of other species. The human polypeptides encoded by SEQ ID NOS:1-800 are likely to be less immunogenic when administered to humans than homologs from other species. Further, on administration to humans, the polypeptides encoded by SEQ ID NOS:1-800 can show greater specificity or can be better regulated by other human proteins than are homologs from other species.

Example 3: Members of Protein Families

After conducting a profile search as described in the specification above, several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 3). Thus the invention encompasses fragments, fusions, and variants of such polynucleotides that retain biological activity associated with the protein family and/or functional domain identified herein.

Table 3 Polynucleotides encoding gene products of a protein family or having a known functional domain(s).

SEQ ID NO:	Biological Activity (Profile hit)	Start	Stop	Dir
24	4 transmembrane segments integral membrane proteins	1218	578	rev
41	4 transmembrane segments integral membrane proteins	1086	413	rev
101	4 transmembrane segments integral membrane proteins	1206	544	rev
157	4 transmembrane segments integral membrane proteins	721	33	rev
341	4 transmembrane segments integral membrane proteins	1253	613	rev
395	4 transmembrane segments integral membrane proteins	530	10	for
395	4 transmembrane segments integral membrane proteins	696	17	for
395	4 transmembrane segments integral membrane proteins	471	39	rev
24	7 transmembrane receptor (Secretin family)	1301	491	rev
41	7 transmembrane receptor (Secretin family)	1309	10	rev
101	7 transmembrane receptor (Secretin family)	1330	296	rev
157	7 transmembrane receptor (Secretin family)	1173	249	rev
291	7 transmembrane receptor (Secretin family)	1400	269	rev
291	7 transmembrane receptor (Secretin family)	712	130	for
305	7 transmembrane receptor (Secretin family)	926	4	for
305	7 transmembrane receptor (Secretin family)	753	55	rev
315	7 transmembrane receptor (Secretin family)	1058	270	rev
341	7 transmembrane receptor (Secretin family)	1265	534	rev
116	Ank repeat	141	218	for
251	Ank repeat	290	207	for
251	Ank repeat	467	387	for
63	ATPases Associated with Various Cellular Activities	543	60	for
116	ATPases Associated with Various Cellular Activities	802	313	for
134	ATPases Associated with Various Cellular Activities	525	57	rev
136	ATPases Associated with Various Cellular Activities	712	163	for
151	ATPases Associated with Various Cellular Activities	719	73	for
151	ATPases Associated with Various Cellular Activities	386	13	for
384	ATPases Associated with Various Cellular Activities	664	140	for
404	ATPases Associated with Various Cellular Activities	704	52	for
374	Basic region plus leucine zipper transcription factors	298	146	for
97	Bromodomain (conserved sequence found in human, Drosophila and yeast proteins.)	230	63	for
136	EF-hand	121	207	for
242	EF-hand	238	155	for
379	EF-hand	212	126	for
308	Eukaryotic aspartyl proteases	1300	461	rev
213	GATA family of transcription factors	720	377	for
367	G-protein alpha subunit	971	467	rev
188	Phorbol esters/diacylglycerol binding	91	177	for
251	Phorbol esters/diacylglycerol binding	133	219	for
202	protein kinase	482	1	rev
202	protein kinase	970	1	rev
315	protein kinase	739	158	for

Table 3 Polynucleotides encoding gene products of a protein family or having a known functional domain(s).

SEQ ID NO:	Biological Activity (Profile hit)	Start	Stop	Dir
315	protein kinase	1023	197	for
367	protein kinase	1046	285	rev
397	protein kinase	511	6	for
256	Protein phosphatase 2C	13	90	for
256	Protein phosphatase 2C	163	86	for
382	Protein Tyrosine Phosphatase	261	2	for
306	SH3 Domain	141	296	for
386	SH3 Domain	359	209	for
169	Trypsin	764	164	rev
188	WD domain, G-beta repeats	480	382	for
188	WD domain, G-beta repeats	206	117	for
335	WD domain, G-beta repeats	3	92	for
23	wnt family of developmental signaling proteins	1151	335	rev
291	wnt family of developmental signaling proteins	779	89	rev
291	wnt family of developmental signaling proteins	1347	382	rev
324	wnt family of developmental signaling proteins	1180	499	rev
330	wnt family of developmental signaling proteins	1180	499	rev
341	wnt family of developmental signaling proteins	1399	560	rev
353	wnt family of developmental signaling proteins	880	49	rev
188	WW/rsp5/WWP domain containing proteins	431	354	for
379	WW/rsp5/WWP domain containing proteins	12	89	for
395	WW/rsp5/WWP domain containing proteins	153	76	for
395	WW/rsp5/WWP domain containing proteins	156	64	for
61	Zinc finger, C2H2 type	254	192	for
306	Zinc finger, C2H2 type	428	367	for
386	Zinc finger, C2H2 type	191	253	for
322	Zinc finger, CCHC class	553	503	for
306	Zinc-binding metalloprotease domain	101	60	rev
395	Zinc-binding metalloprotease domain	28	69	rev

Start and stop indicate the position within the individual sequenes that align with the query sequence having the indicated SEQ ID NO. The direction (Dir) indicates the orientation of the query sequence with respect to the individual sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Some polynucleotides exhibited multiple profile hits because, for example, the particular sequence contains overlapping profile regions, and/or the sequence contains two different functional domains. These profile hits are described in more detail below.

a) Four Transmembrane Integral Membrane Proteins. SEQ ID NOS: 24, 41, 101, 157, 341, and 395 correspond to a sequence encoding a polypeptide that is a member of the 4 transmembrane segments integral membrane protein family (transmembrane 4 family). The transmembrane 4 family of proteins includes a number of evolutionarily-related eukaryotic cell surface antigens (Levy *et al.*, *J. Biol. Chem.*, (1991) *266*:14597; Tomlinson *et al.*, *Eur. J. Immunol.* (1993) *23*:136; Barclay *et al.* The leucocyte antigen factbooks. (1993) Academic Press, London/San Diego). The proteins belonging to this family include: 1) Mammalian antigen CD9 (MIC3), which is involved in platelet activation and aggregation; 2) Mammalian leukocyte antigen CD37, expressed on B lymphocytes; 3) Mammalian leukocyte antigen CD53 (OX-44), which is implicated in growth regulation in hematopoietic cells; 4) Mammalian lysosomal membrane protein CD63 (melanoma-associated antigen ME491; antigen AD1); 5) Mammalian antigen CD81 (cell surface protein TAPA-1), which is implicated in regulation of lymphoma cell growth; 6) Mammalian antigen CD82 (protein R2; antigen C33; Kangai 1 (KAI1)), which associates with CD4 or CD8 and delivers costimulatory signals for the TCR/CD3 pathway; 7) Mammalian antigen CD151 (SFA-1; platelet-endothelial tetraspan antigen 3 (PETA-3)); 8) Mammalian cell surface glycoprotein A15 (TALLA-1; MXS1); 9) Mammalian novel antigen 2 (NAG-2); 10) Human tumor-associated antigen CO-029; 11) Schistosoma mansoni and japonicum 23 Kd surface antigen (SM23 / SJ23).

The members of the 4 transmembrane family share several characteristics. First, they all are apparently type III membrane proteins, which are integral membrane proteins containing an N-terminal membrane-anchoring domain which is not cleaved during biosynthesis and which functions both as a translocation signal and as a membrane anchor. The family members also contain three additional transmembrane regions, at least seven conserved cysteines residues, and are of approximately the same size (218 to 284 residues). These proteins are collectively know as the "transmembrane 4 superfamily" (TM4) because they span plasma membrane four times. A schematic diagram of the domain structure of these proteins is as follows:

```
+-+-----+-------+----+-----+-----+----------------------+-----+----+
| | TMa | Extra | TM2| Cyt | TM3 | Extracellular       | TM4 | Cyt|
+-+-----+-------+----C-----C-----+--------CC---C-----C---+-----C----+
                                          *********
```

where Cyt is the cytoplasmic domain, TMa is the transmembrane anchor; TM2 to TM4 represents transmembrane regions 2 to 4, 'C' are conserved cysteines, and '* 'indicates the position of the consensus pattern. The consensus pattern spans a conserved region including two cysteines located in a short cytoplasmic loop between two transmembrane domains: Consensus pattern: G-x(3)-[LIVMF]-x(2)-[GSA]-[LIVMF](2)-G-C-x-[GA]-[STA]- x(2)-[EG]-x(2)-[CWN]-[LIVM](2).

b) Seven Transmembrane Integral Membrane Proteins. SEQ ID NOS: 24, 41, 101, 157, 291, 305, 315, and 341 correspond to a sequence encoding a polypeptide that is a member of the seven transmembrane receptor family. G-protein coupled receptors (Strosberg, *Eur. J. Biochem.* (1991) *196*:1; Kerlavage, *Curr. Opin. Struct. Biol.* (1991) *1*:394; and Probst *et al.*, *DNA Cell Biol.* (1992) *11*:1; and Savarese *et al.*, *Biochem. J.* (1992) *293*:1) (also called R7G) are an extensive group of hormones, neurotransmitters, odorants and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins. The tertiary structure of these receptors is thought to be highly similar. They have seven hydrophobic regions, each of which most probably spans the membrane. The N-terminus is located on the extracellular side of the membrane and is often glycosylated, while the C-terminus is cytoplasmic and generally phosphorylated. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. Most, but not all of these receptors, lack a signal peptide. The most conserved parts of these proteins are the transmembrane regions and the first two cytoplasmic loops. A conserved acidic-Arg-aromatic triplet is present in the N-terminal extremity of the second cytoplasmic loop (Attwood *et al.*, *Gene* (1991) *98*:153) and could be implicated in the interaction with G proteins.

To detect this widespread family of proteins a pattern is used that contains the conserved triplet and that also spans the major part of the third transmembrane helix. Additional information about the seven transmembrane receptor family, and methods for their identification

and use, is found in U.S. Patent No. 5,759,804. Due in part to their expression on the cell surface and other attractive characteristics, seven transmembrane protein family members are of particular interest as drug targets, as surface antigen markers, and as drug delivery targets (*e.g.*, using antibody-drug complexes and/or use of anti-seven transmembrane protein antibodies as therapeutics in their own right).

c) Ank Repeats. SEQ ID NOS: 116 and 251 represent polynucleotides encoding Ank repeat-containing proteins. The ankyrin motif is a 33 amino acid sequence named after the protein ankyrin which has 24 tandem 33-amino-acid motifs. Ank repeats were originally identified in the cell-cycle-control protein cdc10 (Breeden *et al., Nature* (1987) *329*:651). Proteins containing ankyrin repeats include ankyrin, myotropin, I-kappaB proteins, cell cycle protein cdc10, the Notch receptor (Matsuno *et al., Development* (1997) *124(21)*:4265); G9a (or BAT8) of the class III region of the major histocompatibility complex (Biochem J. 290:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, and SW16. The functions of the ankyrin repeats are compatible with a role in protein-protein interactions (Bork, *Proteins* (1993) *17(4)*:363; Lambert and Bennet, *Eur. J. Biochem.* (1993) *211*:1; Kerr *et al., Current Op. Cell Biol.* (1992) *4*:496; Bennet *et al., J. Biol. Chem.* (1980) *255*:6424).

The 90 kD N-terminal domain of ankyrin contains a series of 24 33-amino-acid ank repeats. (Lux *et al., Nature* (1990) *344*:36-42, Lambert *et al., PNAS USA* (1990) *87*:1730.) The 24 ank repeats form four folded subdomains of 6 repeats each. These four repeat subdomains mediate interactions with at least 7 different families of membrane proteins. Ankyrin contains two separate binding sites for anion exchanger dimers. One site utilizes repeat subdomain two (repeats 7-12) and the other requires both repeat subdomains 3 and 4 (repeats 13-24). Since the anion exchangers exist in dimers, ankyrin binds 4 anion exchangers at the same time. (Michaely and Bennett, *J. Biol. Chem.* (1995) *270(37)*:22050) The repeat motifs are involved in ankyrin interaction with tubulin, spectrin, and other membrane proteins. (Lux *et al., Nature* (1990) *344*:36.)

The Rel/NF-kappaB/Dorsal family of transcription factors have activity that is controlled by sequestration in the cytoplasm in association with inhibitory proteins referred to as I-kappaB. (Gilmore, *Cell* (1990) *62*:841; Nolan and Baltimore, *Curr Opin Genet Dev.* (1992) *2*:211;

Baeuerle, *Biochim Biophys Acta* (1991) *1072*:63; Schmitz *et al.*, *Trends Cell Biol.* (1991) *1*:130.) I-kappaB proteins contain 5 to 8 copies of 33 amino acid ankyrin repeats and certain NF-kappaB/rel proteins are also regulated by cis-acting ankyrin repeat containing domains including p105NF-kappaB which contains a series of ankyrin repeats (Diehl and Hannink, *J. Virol.* (1993) *67(12)*:7161). The I-kappaBs and Cactus (also containing ankyrin repeats) inhibit activators through differential interactions with the Rel-homology domain. The gene family includes proto-oncogenes, thus broadly implicating I-kappaB in the control of both normal gene expression and the aberrant gene expression that makes cells cancerous. (Nolan and Baltimore, *Curr Opin Genet Dev.* (1992) *2(2)*:211-220). In the case of rel/NF-kappaB and pp40/I-kappaBβ, both the ankyrin repeats and the carboxy-terminal domain are required for inhibiting DNA-binding activity and direct association of pp40/I-kappaBβ with rel/NF-kappaB protein. The ankyrin repeats and the carboxy-terminal of pp40/I-kappaBβ (form a structure that associates with the rel homology domain to inhibit DNA binding activity (Inoue *et al.*, *PNAS USA* (1992) *89*:4333).

The 4 ankyrin repeats in the amino terminus of the transcription factor subunit GABPβ are required for its interaction with the GABPα subunit to form a functional high affinity DNA-binding protein. These repeats can be crosslinked to DNA when GABP is bound to its target sequence. (Thompson *et al.*, *Science* (1991) *253*:762; LaMarco *et al.*, *Science* (1991) *253*:789).

Myotrophin, a 12.5 kDa protein having a key role in the initiation of cardiac hypertrophy, comprises ankyrin repeats. The ankyrin repeats are characteristic of a hairpin-like protruding tip followed by a helix-turn-helix motif. The V-shaped helix-turn-helix of the repeats stack sequentially in bundles and are stabilized by compact hydrophobic cores, whereas the protruding tips are less ordered.

d) ATPases Associated with Various Cellular Activities (AAA). SEQ ID NOS: 63, 116, 134, 136, 151, 384, and 404 polynucleotides encoding novel members of the "ATPases Associated with diverse cellular Activities" (AAA) protein family The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids that contains an ATP-binding site (Froehlich *et al.*, *J. Cell Biol.* (1991) *114*:443; Erdmann *et al. Cell* (1991) *64*:499; Peters *et al.*, *EMBO J.* (1990) *9*:1757; Kunau *et al.*, *Biochimie* (1993)

75:209-224; Confalonieri *et al., BioEssays* (1995) *17*:639; http://yeamob.pci.chemie.uni-tuebingen.de/AAA/Description.html). The proteins that belong to this family either contain one or two AAA domains.

Proteins containing two AAA domains include: 1) Mammalian and drosophila NSF (N-ethylmaleimide-sensitive fusion protein) and the fungal homolog, SEC18, which are involved in intracellular transport between the endoplasmic reticulum and Golgi, as well as between different Golgi cisternae; 2) Mammalian transitional endoplasmic reticulum ATPase (previously known as p97 or VCP), which is involved in the transfer of membranes from the endoplasmic reticulum to the golgi apparatus. This ATPase forms a ring-shaped homooligomer composed of six subunits. The yeast homolog, CDC48, plays a role in spindle pole proliferation; 3) Yeast protein PAS1 essential for peroxisome assembly and the related protein PAS1 from Pichia pastoris; 4) Yeast protein AFG2; 5) Sulfolobus acidocaldarius protein SAV and Halobacterium salinarium cdcH, which may be part of a transduction pathway connecting light to cell division.

Proteins containing a single AAA domain include: 1) Escherichia coli and other bacteria ftsH (or hflB) protein. FtsH is an ATP-dependent zinc metallopeptidase that degrades the heat-shock sigma-32 factor, and is an integral membrane protein with a large cytoplasmic C-terminal domain that contain both the AAA and the protease domains; 2) Yeast protein YME1, a protein important for maintaining the integrity of the mitochondrial compartment. YME1 is also a zinc-dependent protease; 3) Yeast protein AFG3 (or YTA10). This protein also contains an AAA domain followed by a zinc-dependent protease domain; 4) Subunits from regulatory complex of the 26S proteasome (Hilt *et al., Trends Biochem. Sci.* (1996) *21*:96), which is involved in the ATP-dependent degradation of ubiquitinated proteins, which subunits include: a) Mammalian 4 and homologs in other higher eukaryotes, in yeast (gene YTA5) and fission yeast (gene mts2); b) Mammalian 6 (TBP7) and homologs in other higher eukaryotes and in yeast (gene YTA2); c) Mammalian subunit 7 (MSS1) and homologs in other higher eukaryotes and in yeast (gene CIM5 or YTA3); d) Mammalian subunit 8 (P45) and homologs in other higher eukaryotes and in yeast (SUG1 or CIM3 or TBY1) and fission yeast (gene let1); e) Other probable subunits include human TBP1, which influences HIV gene expression by interacting with the virus tat transactivator protein, and yeast YTA1 and YTA6; 5) Yeast protein BCS1, a mitochondrial

protein essential for the expression of the Rieske iron-sulfur protein; 6) Yeast protein MSP1, a protein involved in intramitochondrial sorting of proteins; 7) Yeast protein PAS8, and the corresponding proteins PAS5 from Pichia pastoris and PAY4 from Yarrowia lipolytica; 8) Mouse protein SKD1 and its fission yeast homolog (SpAC2G11.06); 9) Caenorhabditis elegans meiotic spindle formation protein mei-1; 10) Yeast protein SAP1' 11) Yeast protein YTA7; and 12) Mycobacterium leprae hypothetical protein A2126A.

In general, the AAA domains in these proteins act as ATP-dependent protein clamps(Confalonieri *et al.* (1995) *BioEssays 17*:639). In addition to the ATP-binding 'A' and 'B' motifs, which are located in the N-terminal half of this domain, there is a highly conserved region located in the central part of the domain which was used in the development of the signature pattern.

e) <u>Basic Region Plus Leucine Zipper Transcription Factors.</u> SEQ ID NO:374 correspond to a polynucleotide encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization. Members of the family include transcription factor AP-1, which binds selectively to enhancer elements in the cis control regions of SV40 and metallothionein IIA. AP-1, also known as c-jun, is the cellular homolog of the avian sarcoma virus 17 (ASV17) oncogene v-jun.

Other members of this protein family include jun-B and jun-D, probable transcription factors that are highly similar to jun/AP-1; the fos protein, a proto-oncogene that forms a non-covalent dimer with c-jun; the fos-related proteins fra-1, and fos B; and mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-3, ATF-4, ATF-5, ATF-6 and LRF-1.

f) <u>Bromodomain.</u> SEQ ID NO:97 corresponds to a polynucleotide encoding a polypeptide having a bromodomain region (Haynes et al., 1992, Nucleic Acids Res. 20:2693-2603, Tamkun et al., 1992, Cell 68:561-572, and Tamkun, 1995, Curr. Opin. Genet. Dev. 5:473-477), which is a conserved region of about 70 amino acids found in the following proteins:

1) Higher eukaryotes transcription initiation factor TFIID 250 Kd subunit (TBP-associated factor p250) (gene CCG1); P250 is associated with the TFIID TATA-box binding protein and seems essential for progression of the G1 phase of the cell cycle. 2) Human RING3, a protein of unknown function encoded in the MHC class II locus; 3) Mammalian CREB-binding protein (CBP), which mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein; 4) Mammalian homologs of brahma, including three brahma-like human: SNF2a(hBRM), SNF2b, and BRG1; 5) Human BS69, a protein that binds to adenovirus E1A and inhibits E1A transactivation; 6) Human peregrin (or Br140).

The bromodomain is thought to be involved in protein-protein interactions and may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation.

g) EF-Hand. SEQ ID NOS:136, 242, and 379 correspond to polynucleotides encoding a novel protein in the family of EF-hand proteins. Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand (Kawasaki *et al.*, *Protein. Prof.* (1995) *2*:305-490). This type of domain consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand).

Proteins known to contain EF-hand regions include: Calmodulin (Ca=4, except in yeast where Ca=3) ("Ca=" indicates approximate number of EF-hand regions); diacylglycerol kinase (EC 2.7.1.107) (DGK) (Ca=2); 2) FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) from mammals (Ca=1); guanylate cyclase activating protein (GCAP) (Ca=3); MIF related proteins 8 (MRP-8 or CFAG) and 14 (MRP-14) (Ca=2); myosin regulatory light chains (Ca=1); oncomodulin (Ca=2); osteonectin (basement membrane protein BM-40) (SPARC); and proteins that contain an "osteonectin" domain (QR1, matrix glycoprotein SC1).

The consensus pattern includes the complete EF-hand loop as well as the first residue which follows the loop and which seem to always be hydrophobic.

h) Eukaryotic Aspartyl Proteases. SEQ ID NO:308 corresponds to a gene encoding a novel eukaryotic aspartyl protease. Aspartyl proteases, known as acid proteases, (EC 3.4.23.-) are a widely distributed family of proteolytic enzymes (Foltmann B., *Essays Biochem.* (1981) *17*:52; Davies D.R., *Annu. Rev. Biophys. Chem.* (1990) *19*:189; Rao J.K.M., *et al.*, *Biochemistry* (1991) *30*:4663) known to exist in vertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Currently known eukaryotic aspartyl proteases include: 1) Vertebrate gastric pepsins A and C (also known as gastricsin); 2) Vertebrate chymosin (rennin), involved in digestion and used for making cheese; 3) Vertebrate lysosomal cathepsins D (EC 3.4.23.5) and E (EC 3.4.23.34); 4) Mammalian renin (EC 3.4.23.15) whose function is to generate angiotensin I from angiotensinogen in the plasma; 5) Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21); and 6) Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases; 7) Yeast barrierpepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves alpha-factor and thus acts as an antagonist of the mating pheromone; and 8) Fission yeast sxa1 which is involved in degrading or processing the mating pheromones.

Most retroviruses and some plant viruses, such as badnaviruses, encode for an aspartyl protease which is an homodimer of a chain of about 95 to 125 amino acids. In most retroviruses, the protease is encoded as a segment of a polyprotein which is cleaved during the maturation process of the virus. It is generally part of the pol polyprotein and, more rarely, of the gag polyprotein. Because the sequence around the two aspartates of eukaryotic aspartyl proteases and around the single active site of the viral proteases is conserved, a single signature pattern can be used to identify members of both groups of proteases.

i) GATA Family of Transcription Factors. SEQ ID NO:213 corresponds to a novel member of the GATA family of transcription factors. The GATA family of transcription factors

are proteins that bind to DNA sites with the consensus sequence (A/T)GATA(A/G), found within the regulatory region of a number of genes. Proteins currently known to belong to this family are: 1) GATA-1 (Trainor, C.D., *et al.*, *Nature* (1990) *343*:92) (also known as Eryf1, GF-1 or NF-E1), which binds to the GATA region of globin genes and other genes expressed in erythroid cells. It is a transcriptional activator which probably serves as a general 'switch' factor for erythroid development; 2) GATA-2 (Lee, M.E., *et al.*, *J. Biol. Chem.* (1991) *266*:16188), a transcriptional activator which regulates endothelin-1 gene expression in endothelial cells; 3) GATA-3 (Ho, I.-C., *et al.*, *EMBO J.* (1991) *10*:1187), a transcriptional activator which binds to the enhancer of the T-cell receptor alpha and delta genes; 4) GATA-4 (Spieth, J., *et al.*, *Mol. Cell. Biol.* (1991) *11*:4651), a transcriptional activator expressed in endodermally derived tissues and heart; 5) Drosophila protein pannier (or DGATAa) (gene pnr) which acts as a repressor of the achaete-scute complex (as-c); 6) Bombyx mori BCFI (Drevet, J.R., *et al.*, *J. Biol. Chem.* (1994) *269*:10660), which regulates the expression of chorion genes; 7) Caenorhabditis elegans elt-1 and elt-2, transcriptional activators of genes containing the GATA region, including vitellogenin genes (Hawkins, M.G., *et al.*, *J. Biol. Chem.* (1995) *270*:14666); 8) Ustilago maydis urbs1 (Voisard, C.P.O., *et al.*, *Mol. Cell. Biol.* (1993) *13*:7091), a protein involved in the repression of the biosynthesis of siderophores; 9) Fission yeast protein GAF2.

All these transcription factors contain a pair of highly similar 'zinc finger' type domains with the consensus sequence C-x2-C-x17-C-x2-C. Some other proteins contain a single zinc finger motif highly related to those of the GATA transcription factors. These proteins are: 1) Drosophila box A-binding factor (ABF) (also known as protein serpent (gene srp)) which may function as a transcriptional activator protein and may play a key role in the organogenesis of the fat body; 2) Emericella nidulans are (Arst, H.N., Jr., *et al.*, *Trends Genet.* (1989) *5*:291) a transcriptional activator which mediates nitrogen metabolite repression; 3) Neurospora crassa nit-2 (Fu, Y.-H., *et al.*, *Mol. Cell. Biol.* (1990) *10*:1056), a transcriptional activator which turns on the expression of genes coding for enzymes required for the use of a variety of secondary nitrogen sources, during conditions of nitrogen limitation; 4) Neurospora crassa white collar proteins 1 and 2 (WC-1 and WC-2), which control expression of light-regulated genes; 5) Saccharomyces cerevisiae DAL81 (or UGA43), a negative nitrogen regulatory protein; 6)

Saccharomyces cerevisiae GLN3, a positive nitrogen regulatory protein; 7) Saccharomyces cerevisiae GAT1; 8) Saccharomyces cerevisiae GZF3.

j) G-Protein Alpha Subunit. SEQ ID NO:367 corresponds to a gene encoding a novel polypeptide of the G-protein alpha subunit family. Guanine nucleotide binding proteins (G-proteins) are a family of membrane-associated proteins that couple extracellularly-activated integral-membrane receptors to intracellular effectors, such as ion channels and enzymes that vary the concentration of second messenger molecules. G-proteins are composed of 3 subunits (alpha, beta and gamma) which, in the resting state, associate as a trimer at the inner face of the plasma membrane. The alpha subunit has a molecule of guanosine diphosphate (GDP) bound to it. Stimulation of the G-protein by an activated receptor leads to its exchange for GTP (guanosine triphosphate). This results in the separation of the alpha from the beta and gamma subunits, which always remain tightly associated as a dimer. Both the alpha and beta-gamma subunits are then able to interact with effectors, either individually or in a cooperative manner. The intrinsic GTPase activity of the alpha subunit hydrolyses the bound GTP to GDP. This returns the alpha subunit to its inactive conformation and allows it to reassociate with the beta-gamma subunit, thus restoring the system to its resting state.

G-protein alpha subunits are 350-400 amino acids in length and have molecular weights in the range 40-45 kDa. Seventeen distinct types of alpha subunit have been identified in mammals. These fall into 4 main groups on the basis of both sequence similarity and function: alpha-s, alpha-q, alpha-i and alpha-12 (Simon *et al.*, *Science* (1993) *252*:802). Many alpha subunits are substrates for ADP-ribosylation by cholera or pertussis toxins. They are often N-terminally acylated, usually with myristate and/or palmitoylate, and these fatty acid modifications are probably important for membrane association and high- affinity interactions with other proteins. The atomic structure of the alpha subunit of the G-protein involved in mammalian vision, transducin, has been elucidated in both GTP- and GDB-bound forms, and shows considerable similarity in both primary and tertiary structure in the nucleotide-binding regions to other guanine nucleotide binding proteins, such as p21-ras and EF-Tu.

k) Phorbol Esters/Diacylglycerol Binding. SEQ ID NO:188 and 251 represent polynucleotides encoding a protein belonging to the family including phorbol

esters/diacylglycerol binding proteins. Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC) (Azzi *et al.*, *Eur. J. Biochem.* (1992) *208*:547). Phorbol esters can directly stimulate PKC. The N-terminal region of PKC, known as C1, has been shown (Ono *et al.*, *Proc. Natl. Acad. Sci. USA* (1989) *86*:4868) to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Such a domain has also been found in, for example, the following proteins.

(1) Diacylglycerol kinase (EC 2.7.1.107) (DGK) (Sakane *et al.*, *Nature* (1990) *344*:345), the enzyme that converts DAG into phosphatidate. It contains two copies of the DAG/PE-binding domain in its N-terminal section. At least five different forms of DGK are known in mammals; and

(2) N-chimaerin, a brain specific protein which shows sequence similarities with the BCR protein at its C-terminal part and contains a single copy of the DAG/PE-binding domain at its N-terminal part. It has been shown (Ahmed *et al.*, *Biochem. J.* (1990) *272*:767, and Ahmed *et al.*, *Biochem. J.* (1991) *280*:233) to be able to bind phorbol esters.

The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain. The signature pattern completely spans the DAG/PE domain. The consensus pattern is: H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)-C-x(2)-C-x(4)-[HD]-x(2)-C-x(5,9)-C. All the C and H are probably involved in binding zinc.

l) Protein Kinase. SEQ ID NOS:202, 315, 367, and 397 represent polynucleotides encoding protein kinases. Protein kinases catalyze phosphorylation of proteins in a variety of pathways, and are implicated in cancer. Eukaryotic protein kinases (Hanks S.K., *et al.*, *FASEB J.* (1995) *9*:576; Hunter T., *Meth. Enzymol.* (1991) *200*:3; Hanks S.K., *et al.*, *Meth. Enzymol.* (1991) *200*:38; Hanks S.K., *Curr. Opin. Struct. Biol.* (1991) *1*:369; Hanks S.K., *et al.*, *Science* (1988) *241*:42) are enzymes that belong to a very extensive family of proteins

which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. Two of the conserved regions are the basis for the signature pattern in the protein kinase profile. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme (Knighton D.R., *et al.*, *Science* (1991) *253*:407). The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the alignment in (Hanks S.K., *et al.*, *FASEB J.* (1995) *9*:576) and covers the entire catalytic domain.

The protein kinase profile also detects receptor guanylate cyclases and 2-5A-dependent ribonucleases. Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed previously. The profile also detects Arabidopsis thaliana kinase-like protein TMKL1 which seems to have lost its catalytic activity.

If a protein analyzed includes the two of the above protein kinase signatures, the probability of it being a protein kinase is close to 100%. Eukaryotic-type protein kinases have also been found in prokaryotes such as Myxococcus xanthus (Munoz-Dorado J., *et al.*, *Cell* (1991) *67*:995) and Yersinia pseudotuberculosis. The patterns shown above has been updated since their publication in (Bairoch A., *et al.*, *Nature* (1988) *331*:22).

m) Protein Phosphatase 2C. SEQ ID NO:256 corresponds to a polynucleotide encoding a novel protein phosphatase 2C (PP2C), which is one of the four major classes of mammalian serine/threonine specific protein phosphatases. PP2C (Wenk *et al.*, *FEBS Lett.* (1992) *297*:135) is a monomeric enzyme of about 42 Kd which shows broad substrate specificity and is dependent on divalent cations (mainly manganese and magnesium) for its activity. Three isozymes are currently known in mammals: PP2C-alpha, -beta and -gamma.

n) Protein Tyrosine Phosphatase. SEQ ID NO:382 represents a polynucleotide encoding a protein tyrosine kinase. Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) (Fischer *et al.*, *Science* (1991) *253*:401; Charbonneau *et al.*, *Annu. Rev. Cell Biol.* (1992) *8*:463;

Trowbridge, *J. Biol. Chem.* (1991) *266*:23517; Tonks *et al.*, *Trends Biochem. Sci.* (1989) *14*:497; and Hunter, *Cell* (1989) *58*:1013) catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s).

Soluble PTPases include PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1-like domain and could act at junctions between the membrane and cytoskeleton; PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes that contain two copies of the SH2 domain at its N-terminal extremity.

Dual specificity PTPases include DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1) which dephosphorylates MAP kinase on both Thr-183 and Tyr-185; and DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues.

Structurally, all known receptor PTPases are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains in their extracellular region. The cytoplasmic region generally contains two copies of the PTPAse domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.

PTPase domains consist of about 300 amino acids. There are two conserved cysteines and the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important. The consensus pattern for PTPases is: [LIVMF]-H-C-x(2)-G-x(3)-[STC]-[STAGP]-x-[LIVMFY]; C is the active site residue.

o) <u>SH3 Domain.</u> SEQ ID NO:306 and 386 represent polynucleotides encoding SH3 domain proteins. The Src homology 3 (SH3) domain is a small protein domain of about 60

amino acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (e.g. Src, Abl, Lck) (Mayer *et al.*, *Nature* (1988) *332*:272). The domain has also been found in a variety of intracellular or membrane-associated proteins (Musacchio *et al.*, *FEBS Lett.* (1992) *307*:55; Pawson *et al.*, *Curr. Biol.* (1993) *3*:434; Mayer *et al.*, *Trends Cell Biol.* (1993) *3*:8; and Pawson *et al.*, *Nature* (1995) *373*:573).

The SH3 domain has a characteristic fold that consists of five or six beta-strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices (Kuriyan *et al.*, *Curr. Opin. Struct. Biol.* (1993) *3*:828). It is believed that SH3 domain-containing proteins mediate assembly of specific protein complexes via binding to proline-rich peptides (Morton *et al.*, *Curr. Biol.* (1994) *4*:615). In general, SH3 domains are found as single copies in a given protein, but there is a significant number of proteins with two SH3 domains and a few with 3 or 4 copies.

SH3 domains have been identified in, for example, protein tyrosine kinases, such as the Src, Abl, Bkt, Csk and ZAP70 families of kinases; mammalian phosphatidylinositol-specific phospholipase C-gamma-1 and -2; mammalian phosphatidyl inositol 3-kinase regulatory p85 subunit; mammalian Ras GTPase-activating protein (GAP); mammalian Vav oncoprotein, a guanine nucleotide exchange factor of the CDC24 family; Drosophila lethal(1)discs large-1 tumor suppressor protein (gene Dlg1); mammalian tight junction protein ZO-1; vertebrate erythrocyte membrane protein p55; Caenorhabditis elegans protein lin-2; rat protein CASK; and mammalian synaptic proteins SAP90/PSD-95, CHAPSYN-110/PSD-93, SAP97/DLG1 and SAP102. Novel SH3-domain containing polypeptides will facilitate elucidation of the role of such proteins in important biological pathways, such as ras activation.

p) <u>Trypsin.</u> SEQ ID NO:169 corresponds to a novel serine protease of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature* (1988) *334*:528). Proteases known to belong to the trypsin family include: 1) Acrosin; 2) Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C; 3) Cathepsin

G; 4) Chymotrypsins; 5) Complement components C1r, C1s, C2, and complement factors B, D and I; 6) Complement-activating component of RA-reactive factor; 7) Cytotoxic cell proteases (granzymes A to H); 8) Duodenase I; 9) Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin).; 10) Enterokinase (EC 3.4.21.9) (enteropeptidase); 11) Hepatocyte growth factor activator; 12) Hepsin; 13) Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin); 14) Plasma kallikrein; 15) Mast cell proteases (MCP) 1 (chymase) to 8; 16) Myeloblastin (proteinase 3) (Wegener's autoantigen); 17) Plasminogen activators (urokinase-type, and tissue-type); 18) Trypsins I, II, III, and IV; 19) Tryptases; 20) Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator; 21) Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab; 22) Apolipoprotein(a); 23) Blood fluke cercarial protease; 24) Drosophila trypsin like proteases: alpha, easter, snake-locus; 25) Drosophila protease stubble (gene sb); and 26) Major mite fecal allergen Der p III. All the above proteins belong to family S1 in the classification of peptidases (Rawlings N.D., *et al.*, *Meth. Enzymol.* (1994) *244*:19; http://www.expasy.ch/cgi-bin/lists?peptidas.txt) and originate from eukaryotic species. It should be noted that bacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns.

q) WD Domain, G-Beta Repeats. SEQ ID NOS:188 and 335 represent novel members of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) *56*:615). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat). Such a repetitive segment has been shown to exist in a number of other

proteins including: human LIS1, a neuronal protein involved in type-1 lissencephaly; and mammalian coatomer beta' subunit (beta'-COP), a component of a cytosolic protein complex that reversibly associates with Golgi membranes to form vesicles that mediate biosynthetic protein transport.

r) wnt Family of Developmental Signaling Proteins. SEQ ID NO: 23, 291, 324, 330, 341, and 353 correspond to novel members of the wnt family of developmental signaling proteins. Wnt-1 (previously known as int-1), the seminal member of this family, (Nusse R., *Trends Genet.* (1988) *4*:291) is a proto-oncogene induced by the integration of the mouse mammary tumor virus. It is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous system (CNS). The sequence of wnt-1 is highly conserved in mammals, fish, and amphibians. Wnt-1 was found to be a member of a large family of related proteins (Nusse R., *et al.*, *Cell* (1992) *69*:1073; McMahon A.P., *Trends Genet.* (1992) *8*:1; Moon R.T., *BioEssays* (1993) *15*:91) that are all thought to be developmental regulators. These proteins are known as wnt-2 (also known as irp), wnt-3, -3A, -4, -5A, -5B, -6, -7A, -7B, -8, -8B, -9 and -10. At least four members of this family are present in Drosophila; one of them, wingless (wg), is implicated in segmentation polarity. All these proteins share the following features characteristics of secretory proteins: a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters. The consensus pattern, which is based upon a highly conserved region including three cysteines, is as follows: C-K-C-H-G-[LIVMT]-S-G-x-C. All sequences known to belong to this family are detected by the provided consensus pattern.

s) Ww/rsp5/WWP Domain-Containing Proteins. SEQ ID NOS:188, 379 , and 395 represent polynucleotides encoding a polypeptide in the family of WW/rsp5/WWP domain-containing proteins. The WW domain (Bork *et al.*, *Trends Biochem. Sci.* (1994) *19*:531; Andre *et al.*, *Biochem. Biophys. Res. Commun.* (1994) *205*:1201; Hofmann *et al.*, *FEBS Lett.* (1995) *358*:153; and Sudol *et al.*, *FEBS Lett.* (1995) *369*:67), also known as rsp5 or WWP), was originally discovered as a short conserved region in a number of unrelated proteins, among them

dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown (Chen *et al.*, *Proc. Natl. Acad. Sci. USA* (1995) *92*:7819) to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. It appears to contain beta-strands grouped around four conserved aromatic positions, generally Trp. The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. It is frequently associated with other domains typical for proteins in signal transduction processes.

Proteins containing the WW domain include:

1. Dystrophin, a multidomain cytoskeletal protein. Its longest alternatively spliced form consists of an N-terminal actin-binding domain, followed by 24 spectrin-like repeats, a cysteine-rich calcium-binding domain and a C-terminal globular domain. Dystrophins form tetramers and is thought to have multiple functions including involvement in membrane stability, transduction of contractile forces to the extracellular environment and organization of membrane specialization. Mutations in the dystrophin gene lead to muscular dystrophy of Duchenne or Becker type. Dystrophin contains one WW domain C-terminal of the spectrin-repeats.

2. Vertebrate YAP protein, which is a substrate of an unknown serine kinase. It binds to the SH3 domain of the Yes oncoprotein via a proline-rich region. This protein appears in alternatively spliced isoforms, containing either one or two WW domains.

3. IQGAP, which is a human GTPase activating protein acting on ras. It contains an N-terminal domain similar to fly muscle mp20 protein and a C-terminal ras GTPase activator domain.

For the sensitive detection of WW domains, the profile spans the whole homology region as well as a pattern.

t) <u>Zinc Finger, C2H2 Type.</u> SEQ ID NO:61, 306, and 386 correspond to polynucleotides encoding novel members of the of the C2H2 type zinc finger protein family. Zinc finger domains (Klug *et al.*, *Trends Biochem. Sci.* (1987) *12*:464; Evans *et al.*, *Cell* (1988) *52*:1; Payre *et al.*, *FEBS Lett.* (1988) *234*:245; Miller *et al.*, *EMBO J.* (1985) *4*:1609; and Berg, *Proc. Natl. Acad. Sci. USA* (1988) *85*:99) are nucleic acid-binding protein structures first identified in the

Xenopus transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino acid residues. Two cysteine or histidine residues are positioned at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides.

Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc-dependent DNA or RNA binding property of some members of this class.

Mammalian proteins having a C2H2 zipper include (number in parenthesis indicates number of zinc finger regions in the protein): basonuclin (6), BCL-6/LAZ-3 (6), erythroid krueppel-like transcription factor (3), transcription factors Sp1 (3), Sp2 (3), Sp3 (3) and Sp(4) 3, transcriptional repressor YY1 (4), Wilms' tumor protein (4), EGR1/Krox24 (3), EGR2/Krox20 (3), EGR3/Pilot (3), EGR4/AT133 (4), Evi-1 (10), GLI1 (5), GLI2 (4+), GLI3 (3+), HIV-EP1/ZNF40 (4), HIV-EP2 (2), KR1 (9+), KR2 (9), KR3 (15+), KR4 (14+), KR5 (11+), HF.12 (6+), REX-1 (4), ZfX (13), ZfY (13), Zfp-35 (18), ZNF7 (15), ZNF8 (7), ZNF35 (10), ZNF42/MZF-1 (13), ZNF43 (22), ZNF46/Kup (2), ZNF76 (7), ZNF91 (36), ZNF133 (3).

In addition to the conserved zinc ligand residues, it has been shown that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld *et al.*, *J. Biomol. Struct. Dyn.* (1993) *11*:557) The best conserved position is found four residues after the second cysteine; it is generally an aromatic or aliphatic residue. The consensus pattern for C2H2 zinc fingers is: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H. The two C's and two H's are zinc ligands.

u) Zinc Finger, CCHC Class. SEQ ID NO:322 corresponds to a polynucleotide encoding a novel member of the zinc finger CCHC family. The CCHC zinc finger protein family to date has been mostly composed of retroviral gag proteins (nucleocapsid). The prototype structure of this family is from HIV. The family also contains members involved in

eukaryotic gene regulation, such as C. elegans GLH-1. The consensus sequence of this family is based upon the common structure of an 18-residue zinc finger.

v) Zinc-Binding Metalloprotease Domain. SEQ ID NO:306 and 395 represent polynucleotides encoding novel members of the zinc-binding metalloprotease domain protein family. The majority of zinc-dependent metallopeptidases (with the notable exception of the carboxypeptidases) share a common pattern of primary structure (Jongeneel *et al.*, *FEBS Lett.* (1989) *242*:211; Murphy *et al.*, *FEBS Lett.* (1991) *289*:4; and Bode *et al.*, *Zoology* (1996) *99*:237) in the part of their sequence involved in the binding of zinc, and can be grouped together as a superfamily, known as the metzincins, on the basis of this sequence similarity. Examples of these proteins include: 1) Angiotensin-converting enzyme (EC 3.4.15.1) (dipeptidyl carboxypeptidase I) (ACE), the enzyme responsible for hydrolyzing angiotensin I to angiotensin II. 2) Mammalian extracellular matrix metalloproteinases (known as matrixins) (Woessner, *FASEB J.* (1991) *5*:2145): MMP-1 (EC 3.4.24.7) (interstitial collagenase), MMP-2 (EC 3.4.24.24) (72 Kd gelatinase), MMP-9 (EC 3.4.24.35) (92 Kd gelatinase), MMP-7 (EC 3.4.24.23) (matrylisin), MMP-8 (EC 3.4.24.34) (neutrophil collagenase), MMP-3 (EC 3.4.24.17) (stromelysin-1), MMP-10 (EC 3.4.24.22) (stromelysin-2), and MMP-11 (stromelysin-3), MMP-12 (EC 3.4.24.65) (macrophage metalloelastase). 3) Endothelin-converting enzyme 1 (EC 3.4.24.71) (ECE-1), which processes the precursor of endothelin to release the active peptide.

Example 4: Differential Expression of Polynucleotides of the Invention : Description of Libraries and Detection of Differential Expression

The relative expression levels of the polynucleotides of the invention was assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 4 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepared the cDNA library, the "nickname" of the library that is used in the tables below (in quotes), and the approximate number of clones in the library.

Table 4 Description of cDNA Libraries

Library (lib #)	Description	Number of Clones in this Clustering
1	Km12 L4 Human Colon Cell Line, High Metastatic Potential (derived from Km12C) "High Colon"	307133
2	Km12C Human Colon Cell Line, Low Metastatic Potential "Low Colon"	284755
3	MDA-MB-231 Human Breast Cancer Cell Line, High Metastatic Potential; micro-metastases in lung "High Breast"	326937
4	MCF7 Human Breast Cancer Cell, Non Metastatic "Low Breast"	318979
8	MV-522 Human Lung Cancer Cell Line, High Metastatic Potential "High Lung"	223620
9	UCP-3 Human Lung Cancer Cell Line, Low Metastatic Potential "Low Lung"	312503
12	Human microvascular endothelial cells (HMEC) – Untreated PCR (OligodT) cDNA library	41938
13	Human microvascular endothelial cells (HMEC) – bFGF treated PCR (OligodT) cDNA library	42100
14	Human microvascular endothelial cells (HMEC) – VEGF treated PCR (OligodT) cDNA library	42825
15	Normal Colon – UC#2 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	34285
16	Colon Tumor – UC#2 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	35625
17	Liver Metastasis from Colon Tumor of UC#2 Patient	

Library (lib #)	Description	Number of Clones in this Clustering
	PCR (OligodT) cDNA library "High Colon Metastasis Tissue"	36984
18	Normal Colon – UC#3 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	36216
19	Colon Tumor – UC#3 Patient PCR (OligodT) cDNA library "High Colon Tumor Tissue"	41388
20	Liver Metastasis from Colon Tumor of UC#3 Patient PCR (OligodT) cDNA library "High Colon Metastasis Tissue".	30956

The KM12L4 and KM12C cell lines are described in Example 1 above. The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, *e.g.*, Chandrasekaran *et al.*, *Cancer Res.* (1979) *39*:870 (MDA-MB-231 and MCF-7); Gastpar *et al.*, *J Med Chem* (1998) *41*:4965 (MDA-MB-231 and MCF-7); Ranson *et al.*, *Br J Cancer* (1998) *77*:1586 (MDA-MB-231 and MCF-7); Kuang *et al.*, *Nucleic Acids Res* (1998) *26*:1116 (MDA-MB-231 and MCF-7); Varki *et al.*, *Int J Cancer* (1987) *40*:46 (UCP-3); Varki *et al.*, *Tumour Biol.* (1990) *11*:327; (MV-522 and UCP-3); Varki *et al.*, *Anticancer Res.* (1990) *10*:637; (MV-522); Kelner *et al.*, *Anticancer Res* (1995) *15*:867 (MV-522); and Zhang *et al.*, *Anticancer Drugs* (1997) *8*:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3).

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters.

The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac *et al.*, *Genomics* (1996) *37*(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought together computationally. These groups of clones are termed "clusters". Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the "purity" of each cluster can be controlled. For example, artifacts of clustering may occur in computational clustering just as artifacts can occur in "wet-lab" screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1st), and the determining the number of cDNA clones corresponding to the selected cluster in the second library (Clones in 2nd). Differential expression of the selected cluster in the first library relative to the second library is expressed as a "ratio" of percent expression between the two libraries. In general, the "ratio" is calculated by: 1) calculating the percent expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the "number of

clones" corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the "depth" of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5 , where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar, Biostatistical Analysis, Prentice Hall, Inc., USA, "Differences between Proportions," pp 296-298 (1974).

Tables 5 to 7 (inserted before the claims) show the number of clones in each of the above libraries that were analyzed for differential expression. Examples of differentially expressed polynucleotides of particular interest are described in more detail below.

Table 5

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00001340B:A06	17062	3	0	0	0	0	0
M00001340D:F10	11589	2	2	1	3	3	8
M00001341A:E12	4443	10	6	2	6	3	11
M00001342B:E06	39805	2	0	0	0	1	0
M00001343C:F10	2790	7	15	13	14	6	0
M00001343D:H07	23255	3	0	1	1	0	0
M00001345A:E01	6420	8	0	2	0	1	0
M00001346A:F09	5007	4	8	3	6	2	6
M00001346D:E03	6806	5	2	1	2	0	3
M00001346D:G06	5779	5	4	3	4	0	0
M00001346D:G06	5779	5	4	3	4	0	0
M00001347A:B10	13576	5	0	0	0	12	11
M00001348B:B04	16927	4	0	0	2	0	0
M00001348B:G06	16985	4	0	0	0	0	0
M00001349B:B08	3584	5	11	5	0	0	2
M00001350A:H01	7187	5	3	1	0	1	0
M00001351B:A08	3162	10	14	1	6	6	5
M00001351B:A08	3162	10	14	1	6	6	5
M00001352A:E02	16245	4	0	0	0	0	0
M00001353A:G12	8078	4	3	1	0	1	0
M00001353D:D10	14929	4	0	0	1	23	16
M00001355B:G10	14391	3	1	0	0	0	0

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00001357D:D11	4059	8	6	8	16	0	1
M00001361A:A05	4141	5	2	10	16	4	27
M00001361D:F08	2379	26	13	4	2	2	3
M00001362B:D10	5622	7	4	2	13	1	2
M00001362C:H11	945	9	21	2	1	0	0
M00001365C:C10	40132	2	0	0	0	3	0
M00001370A:C09	6867	7	3	0	0	0	0
M00001371C:E09	7172	3	5	1	2	0	1
M00001376B:G06	17732	1	3	5	0	1	4
M00001378B:B02	39833	2	0	0	0	0	0
M00001379A:A05	1334	27	38	35	28	3	0
M00001380D:B09	39886	2	0	0	0	0	0
M00001382C:A02	22979	2	1	0	0	0	0
M00001383A:C03	39648	2	0	0	0	0	0
M00001383A:C03	39648	2	0	0	0	0	0
M00001386C:B12	5178	5	5	4	2	5	2
M00001387A:C05	2464	5	19	25	16	1	0
M00001387B:G03	7587	6	2	1	0	0	0
M00001388D:G05	5832	10	3	0	1	5	0
M00001389A:C08	16269	3	0	0	0	1	1
M00001394A:F01	6583	2	7	3	2	0	0
M00001395A:C03	4016	5	14	0	6	0	0
M00001396A:C03	4009	6	4	13	5	4	10
M00001402A:E08	39563	2	0	0	0	0	0
M00001407B:D11	5556	8	1	5	0	2	0
M00001409C:D12	9577	5	2	0	1	11	12
M00001410A:D07	7005	8	2	0	0	0	0
M00001412B:B10	8551	4	4	0	3	0	0
M00001415A:H06	13538	5	0	0	0	9	1
M00001416A:H01	7674	5	2	0	5	0	0
M00001416B:H11	8847	4	1	3	0	6	1
M00001417A:E02	36393	2	0	0	1	0	0
M00001418B:F03	9952	4	2	1	1	0	0
M00001418D:B06	8526	3	2	1	5	1	0
M00001421C:F01	9577	5	2	0	1	11	12
M00001423B:E07	15066	4	0	0	0	0	0
M00001424B:G09	10470	5	1	0	2	0	1
M00001425B:H08	22195	3	0	0	0	0	0
M00001426D:C08	4261	4	9	7	9	12	15
M00001428A:H10	84182	1	0	0	0	0	0
M00001429A:H04	2797	15	11	18	16	1	14
M00001429B:A11	4635	7	9	2	0	0	0
M00001429D:D07	40392	2	0	1	8	12	16
M00001439C:F08	40054	1	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00001442C:D07	16731	3	1	0	0	0	0
M00001445A:F05	13532	3	2	1	0	1	2
M00001446A:F05	7801	5	2	4	6	1	0
M00001447A:G03	10717	7	2	0	5	8	0
M00001448D:C09	8	1850	2127	1703	3133	1355	122
M00001448D:H01	36313	2	0	0	0	1	30
M00001449A:A12	5857	6	2	3	4	0	0
M00001449A:B12	41633	1	1	0	0	0	0
M00001449A:D12	3681	12	5	10	1	2	5
M00001449A:G10	36535	2	0	0	0	0	0
M00001449C:D06	86110	1	0	0	0	0	0
M00001450A:A02	39304	2	0	0	0	0	0
M00001450A:A11	32663	1	1	0	0	0	0
M00001450A:B12	82498	1	0	0	0	0	0
M00001450A:D08	27250	2	0	0	0	0	0
M00001452A:B04	84328	1	0	0	0	0	0
M00001452A:B12	86859	1	0	0	0	0	0
M00001452A:D08	1120	44	41	5	11	5	0
M00001452A:F05	85064	1	0	0	0	0	0
M00001452C:B06	16970	4	0	0	0	3	4
M00001453A:E11	16130	3	1	0	0	0	1
M00001453C:F06	16653	3	1	0	0	0	0
M00001454A:A09	83103	1	0	0	0	0	0
M00001454B:C12	7005	8	2	0	0	0	0
M00001454D:G03	689	58	95	17	36	66	95
M00001455A:E09	13238	4	1	0	0	0	0
M00001455B:E12	13072	4	1	0	0	0	0
M00001455D:F09	9283	4	1	0	1	0	1
M00001455D:F09	9283	4	1	0	1	0	1
M00001460A:F06	2448	23	22	2	3	3	1
M00001460A:F12	39498	2	0	0	0	0	0
M00001461A:D06	1531	20	23	32	17	14	14
M00001463C:B11	19	1415	1203	1364	525	479	774
M00001465A:B11	10145	2	0	2	0	0	0
M00001466A:E07	4275	11	2	5	0	4	2
M00001467A:B07	38759	2	0	0	0	1	1
M00001467A:D04	39508	2	0	0	0	0	0
M00001467A:D08	16283	3	0	0	0	0	0
M00001467A:D08	16283	3	0	0	0	0	0
M00001467A:E10	39442	2	0	0	0	0	0
M00001468A:F05	7589	6	2	1	1	1	0
M00001469A:C10	12081	4	0	0	0	0	0
M00001469A:H12	19105	2	0	2	0	1	0
M00001470A:B10	1037	53	48	4	22	0	0

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00001470A:C04	39425	2	0	0	0	0	0
M00001471A:B01	39478	2	0	0	0	0	0
M00001481D:A05	7985	3	1	4	0	1	0
M00001490B:C04	18699	2	1	0	0	0	3
M00001494D:F06	7206	4	3	3	1	2	0
M00001497A:G02	2623	12	4	31	4	6	1
M00001499B:A11	10539	2	1	1	0	1	0
M00001500A:C05	5336	9	2	4	8	3	15
M00001500A:E11	2623	12	4	31	4	6	1
M00001500C:E04	9443	4	2	1	1	0	0
M00001501D:C02	9685	3	2	0	7	2	3
M00001504C:A07	10185	5	1	0	0	2	4
M00001504C:H06	6974	7	3	0	1	0	0
M00001504D:G06	6420	8	0	2	0	1	0
M00001507A:H05	39168	2	0	0	0	0	0
M00001511A:H06	39412	2	0	0	0	0	0
M00001512A:A09	39186	2	0	0	0	0	0
M00001512D:G09	3956	9	9	5	2	0	0
M00001513A:B06	4568	10	4	0	9	2	0
M00001513C:E08	14364	1	0	0	0	0	0
M00001514C:D11	40044	2	0	0	0	0	0
M00001517A:B07	4313	13	6	1	0	1	0
M00001518C:B11	8952	3	4	0	4	2	0
M00001528A:C04	7337	4	4	3	16	12	21
M00001528A:F09	18957	3	0	0	0	0	0
M00001528B:H04	8358	3	3	2	0	0	0
M00001531A:D01	38085	2	0	0	0	0	0
M00001532B:A06	3990	6	12	4	1	3	1
M00001533A:C11	2428	14	14	13	9	2	19
M00001534A:C04	16921	4	0	0	1	2	1
M00001534A:D09	5097	6	5	1	1	3	2
M00001534A:F09	5321	11	7	1	5	10	26
M00001534C:A01	4119	9	4	2	2	5	3
M00001535A:B01	7665	3	1	5	0	0	0
M00001535A:C06	20212	2	0	1	1	0	0
M00001535A:F10	39423	2	0	0	0	0	0
M00001536A:B07	2696	23	11	9	18	10	21
M00001536A:C08	39392	2	0	0	0	0	0
M00001537A:F12	39420	2	0	0	0	0	0
M00001537B:G07	3389	4	11	13	2	0	0
M00001540A:D06	8286	6	1	0	3	4	0
M00001541A:D02	3765	19	6	0	0	0	0
M00001541A:F07	22085	3	0	0	0	0	1
M00001541A:H03	39174	2	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00001542A:A09	22113	3	0	0	0	0	0
M00001542A:E06	39453	2	0	0	0	0	0
M00001544A:E03	12170	2	1	2	0	0	0
M00001544A:G02	19829	2	0	1	0	0	0
M00001544B:B07	6974	7	3	0	1	0	0
M00001545A:C03	19255	2	0	0	0	0	0
M00001545A:D08	13864	3	0	2	1	2	4
M00001546A:G11	1267	43	55	5	0	0	0
M00001548A:E10	5892	5	1	4	4	1	3
M00001548A:H09	1058	40	44	37	47	39	59
M00001549A:B02	4015	10	5	8	15	2	0
M00001549A:D08	10944	3	0	3	1	0	7
M00001549B:F06	4193	12	7	2	2	0	1
M00001549C:E06	16347	4	0	0	0	0	0
M00001550A:A03	7239	5	2	1	0	2	0
M00001550A:G01	5175	8	1	3	2	0	0
M00001551A:B10	6268	6	4	3	18	5	0
M00001551A:F05	39180	2	0	0	0	0	0
M00001551A:G06	22390	2	1	0	0	0	1
M00001551C:G09	3266	12	14	0	1	0	6
M00001552A:B12	307	73	60	196	75	79	27
M00001552A:D11	39458	2	0	0	0	0	0
M00001552B:D04	5708	5	4	4	3	1	4
M00001553A:H06	8298	4	3	1	3	0	0
M00001553B:F12	4573	5	7	2	5	0	1
M00001553D:D10	22814	3	0	0	0	0	0
M00001555A:B02	39539	2	0	0	0	1	0
M00001555A:C01	39195	2	0	0	0	0	0
M00001555D:G10	4561	8	4	4	8	0	0
M00001556A:C09	9244	2	0	3	2	10	17
M00001556A:F11	1577	12	40	25	3	4	0
M00001556A:H01	15855	2	1	1	2	12	213
M00001556B:C08	4386	7	8	3	1	3	21
M00001556B:G02	11294	4	0	2	0	0	1
M00001557A:D02	7065	5	3	2	1	0	0
M00001557A:D02	7065	5	3	2	1	0	0
M00001557A:F01	9635	3	0	2	1	0	0
M00001557A:F03	39490	2	0	0	0	1	0
M00001557B:H10	5192	8	5	0	5	0	0
M00001557D:D09	8761	3	4	0	1	0	1
M00001558B:H11	7514	5	3	0	0	0	0
M00001560D:F10	6558	4	3	4	0	0	5
M00001561A:C05	39486	2	0	0	0	0	0
M00001563B:F06	102	289	233	278	116	123	184

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00001564A:B12	5053	11	4	2	2	1	1
M00001571C:H06	5749	4	1	9	0	0	0
M00001578B:E04	23001	2	1	0	2	0	0
M00001579D:C03	6539	8	3	0	0	0	1
M00001583D:A10	6293	3	5	2	6	0	0
M00001586C:C05	4623	3	4	12	2	1	1
M00001587A:B11	39380	2	0	0	0	0	0
M00001594B:H04	260	189	188	27	2	15	0
M00001597C:H02	4837	6	2	10	0	3	1
M00001597D:C05	10470	5	1	0	2	0	1
M00001598A:G03	16999	4	0	0	0	0	0
M00001601A:D08	22794	2	0	0	0	0	0
M00001604A:B10	1399	49	27	19	7	10	23
M00001604A:F05	39391	2	0	0	0	0	0
M00001607A:E11	11465	5	0	0	0	0	0
M00001608A:B03	7802	5	4	0	1	0	0
M00001608B:E03	22155	3	0	0	0	0	0
M00001614C:F10	13157	4	1	0	3	1	0
M00001617C:E02	17004	4	0	1	0	1	0
M00001619C:F12	40314	2	0	0	0	1	0
M00001621C:C08	40044	2	0	0	0	0	0
M00001623D:F10	13913	2	1	2	0	0	1
M00001624A:B06	3277	10	11	8	3	5	1
M00001624C:F01	4309	4	13	3	10	0	0
M00001630B:H09	5214	10	2	2	2	4	3
M00001644C:B07	39171	2	0	0	0	0	0
M00001645A:C12	19267	2	0	0	0	0	1
M00001648C:A01	4665	5	9	0	0	0	0
M00001657D:C03	23201	3	0	0	0	3	0
M00001657D:F08	76760	1	0	2	2	0	5
M00001662C:A09	23218	3	0	0	0	0	0
M00001663A:E04	35702	2	0	0	0	0	0
M00001669B:F02	6468	4	3	3	8	1	0
M00001670C:H02	14367	3	0	0	0	0	0
M00001673C:H02	7015	6	3	1	2	1	1
M00001675A:C09	8773	4	1	4	4	4	6
M00001676B:F05	11460	4	2	0	0	0	0
M00001677C:E10	14627	1	2	1	0	1	0
M00001677D:A07	7570	5	3	0	0	0	0
M00001678D:F12	4416	9	5	2	6	1	3
M00001679A:A06	6660	7	0	4	2	1	0
M00001679A:F10	26875	1	0	0	0	1	0
M00001679B:F01	6298	2	4	5	3	1	0
M00001679C:F01	78091	1	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00001679D:D03	10751	3	2	0	1	0	1
M00001679D:D03	10751	3	2	0	1	0	1
M00001680D:F08	10539	2	1	1	0	1	0
M00001682C:B12	17055	4	0	0	0	0	0
M00001686A:E06	4622	7	6	4	2	3	0
M00001688C:F09	5382	6	2	6	2	0	3
M00001693C:G01	4393	10	6	2	4	1	1
M00001716D:H05	67252	1	0	0	1	0	0
M00003741D:C09	40108	2	0	0	0	0	0
M00003747D:C05	11476	6	0	0	0	0	0
M00003759B:B09	697	76	52	30	72	21	30
M00003762C:B08	17076	4	0	0	0	0	0
M00003763A:F06	3108	14	11	7	5	0	1
M00003774C:A03	67907	1	0	0	0	0	0
M00003796C:D05	5619	3	5	3	3	0	4
M00003826B:A06	11350	3	3	0	0	1	0
M00003833A:E05	21877	2	1	0	0	0	1
M00003837D:A01	7899	5	4	0	2	1	0
M00003839A:D08	7798	5	2	2	0	0	1
M00003844C:B11	6539	8	3	0	0	0	1
M00003846B:D06	6874	6	3	0	0	0	0
M00003851B:D10	13595	4	0	1	0	0	1
M00003853A:D04	5619	3	5	3	3	0	4
M00003853A:F12	10515	5	1	0	1	1	2
M00003856B:C02	4622	7	6	4	2	3	0
M00003857A:G10	3389	4	11	13	2	0	0
M00003857A:H03	4718	4	5	5	2	4	6
M00003871C:E02	4573	5	7	2	5	0	1
M00003875B:F04	12977	5	0	0	0	0	0
M00003875B:F04	12977	5	0	0	0	0	0
M00003875C:G07	8479	4	3	1	1	2	4
M00003876D:E12	7798	5	2	2	0	0	1
M00003879B:C11	5345	7	1	7	4	6	27
M00003879B:D10	31587	1	1	0	0	1	0
M00003879D:A02	14507	3	1	0	0	3	1
M00003885C:A02	13576	5	0	0	0	12	11
M00003885C:A02	13576	5	0	0	0	12	11
M00003906C:E10	9285	4	3	0	0	1	2
M00003907D:A09	39809	1	0	0	0	2	1
M00003907D:H04	16317	3	0	0	0	0	0
M00003909D:C03	8672	4	4	0	0	0	0
M00003912B:D01	12532	4	1	0	1	0	1
M00003914C:F05	3900	9	6	8	1	7	13
M00003922A:E06	23255	3	0	1	1	0	0

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00003958A:H02	18957	3	0	0	0	0	0
M00003958A:H02	18957	3	0	0	0	0	0
M00003958C:G10	40455	2	0	0	0	0	0
M00003958C:G10	40455	2	0	0	0	0	0
M00003968B:F06	24488	2	0	1	4	0	0
M00003970C:B09	40122	2	0	0	0	0	0
M00003974D:E07	23210	3	0	0	0	0	0
M00003974D:H02	23358	3	0	0	0	1	0
M00003975A:G11	12439	4	0	0	0	0	0
M00003978B:G05	5693	7	4	1	3	1	1
M00003981A:E10	3430	9	10	7	3	0	0
M00003982C:C02	2433	10	13	21	18	8	8
M00003983A:A05	9105	5	1	1	1	0	0
M00004028D:A06	6124	4	8	1	9	1	0
M00004028D:C05	40073	2	0	1	0	0	1
M00004031A:A12	9061	5	2	0	0	0	0
M00004031A:A12	9061	5	2	0	0	0	0
M00004035C:A07	37285	2	0	0	1	0	1
M00004035D:B06	17036	4	0	0	0	0	0
M00004059A:D06	5417	10	4	0	9	2	0
M00004068B:A01	3706	7	14	4	22	1	0
M00004072B:B05	17036	4	0	0	0	0	0
M00004081C:D10	15069	3	0	0	1	0	0
M00004081C:D12	14391	3	1	0	0	0	0
M00004086D:G06	9285	4	3	0	0	1	2
M00004087D:A01	6880	2	6	1	1	0	0
M00004093D:B12	5325	5	5	2	0	2	1
M00004093D:B12	5325	5	5	2	0	2	1
M00004105C:A04	7221	5	2	2	2	0	0
M00004108A:E06	4937	4	9	3	1	3	1
M00004111D:A08	6874	6	3	0	0	0	0
M00004114C:F11	13183	2	3	0	7	0	1
M00004138B:H02	13272	3	2	0	3	0	0
M00004146C:C11	5257	2	8	5	5	5	25
M00004151D:B08	16977	4	0	0	0	0	0
M00004157C:A09	6455	3	1	6	0	0	0
M00004169C:C12	5319	6	2	8	2	2	3
M00004171D:B03	4908	6	7	2	2	2	0
M00004172C:D08	11494	4	0	0	0	0	0
M00004183C:D07	16392	3	0	0	0	0	0
M00004185C:C03	11443	5	1	0	0	0	0
M00004197D:H01	8210	2	6	0	0	0	0
M00004203B:C12	14311	4	0	0	0	1	2
M00004212B:C07	2379	26	13	4	2	2	3

Clone Name	Cluster ID	Clones in Lib1	Clones in Lib2	Clones in Lib3	Clones in Lib4	Clones in Lib8	Clones in Lib9
M00004214C:H05	11451	3	2	1	2	1	1
M00004223A:G10	16918	4	0	0	0	0	0
M00004223B:D09	7899	5	4	0	2	1	0
M00004223D:E04	12971	4	0	0	0	1	0
M00004229B:F08	6455	3	1	6	0	0	0
M00004230B:C07	7212	3	5	2	1	3	0
M00004269D:D06	4905	7	6	3	1	3	1
M00004275C:C11	16914	3	0	0	1	0	0
M00004283B:A04	14286	3	1	0	1	1	1
M00004285B:E08	56020	1	0	0	0	0	0
M00004295D:F12	16921	4	0	0	1	2	1
M00004296C:H07	13046	4	1	0	1	0	0
M00004307C:A06	9457	2	0	5	0	3	0
M00004312A:G03	26295	2	0	0	0	0	0
M00004318C:D10	21847	2	1	0	0	0	0
M00004372A:A03	2030	13	10	32	4	0	0
M00004377C:F05	2102	12	20	23	21	6	5

Table 6

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00001340B:A06	17062	0	0	0	0	0	0
M00001340D:F10	11589	0	0	0	0	0	0
M00001341A:E12	4443	0	0	0	1	0	0
M00001342B:E06	39805	0	0	0	0	0	0
M00001343C:F10	2790	0	0	0	0	0	0
M00001343D:H07	23255	0	0	0	0	0	0
M00001345A:E01	6420	0	0	0	0	0	0
M00001346A:F09	5007	0	0	0	0	0	0
M00001346D:E03	6806	0	0	0	0	0	0
M00001346D:G06	5779	0	0	0	0	0	0
M00001346D:G06	5779	0	0	0	0	0	0
M00001347A:B10	13576	0	0	0	0	0	0
M00001348B:B04	16927	0	0	0	0	0	0
M00001348B:G06	16985	0	0	0	0	0	0
M00001349B:B08	3584	0	0	0	0	0	0
M00001350A:H01	7187	0	0	0	0	0	0
M00001351B:A08	3162	0	1	0	0	1	0
M00001351B:A08	3162	0	1	0	0	1	0
M00001352A:E02	16245	0	0	0	0	0	0
M00001353A:G12	8078	0	0	0	0	0	0
M00001353D:D10	14929	0	3	1	0	5	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00001355B:G10	14391	0	0	0	0	0	0
M00001357D:D11	4059	0	0	0	0	0	0
M00001361A:A05	4141	0	0	0	0	0	0
M00001361D:F08	2379	0	0	0	0	0	0
M00001362B:D10	5622	0	0	0	0	0	0
M00001362C:H11	945	0	0	0	0	0	1
M00001365C:C10	40132	0	0	0	0	0	0
M00001370A:C09	6867	0	0	0	0	0	0
M00001371C:E09	7172	0	0	0	0	0	0
M00001376B:G06	17732	0	0	0	0	0	1
M00001378B:B02	39833	0	0	0	0	0	0
M00001379A:A05	1334	0	0	0	0	0	1
M00001380D:B09	39886	0	0	0	0	0	0
M00001382C:A02	22979	0	0	0	0	0	0
M00001383A:C03	39648	0	0	0	0	0	0
M00001383A:C03	39648	0	0	0	0	0	0
M00001386C:B12	5178	0	0	0	0	0	0
M00001387A:C05	2464	0	0	0	0	0	0
M00001387B:G03	7587	0	0	0	0	0	0
M00001388D:G05	5832	0	0	0	0	0	0
M00001389A:C08	16269	0	1	0	0	0	0
M00001394A:F01	6583	1	4	1	0	0	0
M00001395A:C03	4016	0	0	0	0	0	0
M00001396A:C03	4009	0	0	0	0	0	0
M00001402A:E08	39563	0	0	0	0	0	0
M00001407B:D11	5556	0	0	0	0	0	0
M00001409C:D12	9577	0	0	0	0	0	0
M00001410A:D07	7005	0	0	0	0	0	0
M00001412B:B10	8551	0	0	0	0	0	0
M00001415A:H06	13538	0	0	0	0	0	0
M00001416A:H01	7674	0	0	0	0	0	0
M00001416B:H11	8847	0	0	0	0	0	0
M00001417A:E02	36393	0	0	0	0	0	0
M00001418B:F03	9952	0	0	0	0	0	0
M00001418D:B06	8526	0	0	0	0	0	0
M00001421C:F01	9577	0	0	0	0	0	0
M00001423B:E07	15066	0	0	0	0	0	0
M00001424B:G09	10470	0	0	0	0	0	0
M00001425B:H08	22195	0	0	0	0	0	0
M00001426D:C08	4261	0	0	1	0	0	1
M00001428A:H10	84182	0	0	0	0	0	0
M00001429A:H04	2797	0	0	0	0	0	0
M00001429B:A11	4635	0	0	0	0	0	0
M00001429D:D07	40392	0	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00001439C:F08	40054	0	0	0	0	0	0
M00001442C:D07	16731	0	0	0	0	0	0
M00001445A:F05	13532	0	0	0	0	0	0
M00001446A:F05	7801	0	0	0	0	0	0
M00001447A:G03	10717	0	0	0	0	0	0
M00001448D:C09	8	1	6	6	1	14	1
M00001448D:H01	36313	0	3	0	0	3	0
M00001449A:A12	5857	0	0	0	0	0	0
M00001449A:B12	41633	0	0	0	0	0	0
M00001449A:D12	3681	0	0	0	0	0	0
M00001449A:G10	36535	0	0	0	0	0	0
M00001449C:D06	86110	0	0	0	0	0	0
M00001450A:A02	39304	0	0	0	0	0	0
M00001450A:A11	32663	0	0	0	0	0	0
M00001450A:B12	82498	0	0	0	0	0	0
M00001450A:D08	27250	0	0	0	0	0	0
M00001452A:B04	84328	0	0	0	0	0	0
M00001452A:B12	86859	0	0	0	0	0	0
M00001452A:D08	1120	0	0	0	0	0	0
M00001452A:F05	85064	0	0	0	0	0	0
M00001452C:B06	16970	0	0	2	0	1	0
M00001453A:E11	16130	0	0	0	0	0	0
M00001453C:F06	16653	0	0	0	0	0	0
M00001454A:A09	83103	0	0	0	0	0	0
M00001454B:C12	7005	0	0	0	0	0	0
M00001454D:G03	689	0	2	2	0	4	2
M00001455A:E09	13238	0	0	0	0	0	0
M00001455B:E12	13072	0	0	0	0	0	0
M00001455D:F09	9283	0	0	0	0	0	0
M00001455D:F09	9283	0	0	0	0	0	0
M00001460A:F06	2448	0	0	0	0	0	0
M00001460A:F12	39498	0	0	0	0	0	0
M00001461A:D06	1531	0	0	0	0	0	0
M00001463C:B11	19	2	13	13	0	69	10
M00001465A:B11	10145	0	0	0	0	0	0
M00001466A:E07	4275	0	0	0	0	0	0
M00001467A:B07	38759	0	0	0	0	0	0
M00001467A:D04	39508	0	0	0	0	0	0
M00001467A:D08	16283	0	0	0	0	0	0
M00001467A:D08	16283	0	0	0	0	0	0
M00001467A:E10	39442	0	0	0	0	0	0
M00001468A:F05	7589	0	0	0	0	0	0
M00001469A:C10	12081	0	0	0	0	0	0
M00001469A:H12	19105	0	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00001470A:B10	1037	0	0	0	0	0	0
M00001470A:C04	39425	0	0	0	0	0	0
M00001471A:B01	39478	0	0	0	0	0	0
M00001481D:A05	7985	0	0	0	0	0	0
M00001490B:C04	18699	0	0	0	0	0	0
M00001494D:F06	7206	0	0	0	0	0	0
M00001497A:G02	2623	0	0	0	0	0	0
M00001499B:A11	10539	0	0	0	0	0	0
M00001500A:C05	5336	0	0	0	0	0	0
M00001500A:E11	2623	0	0	0	0	0	0
M00001500C:E04	9443	0	0	0	0	0	0
M00001501D:C02	9685	0	0	0	0	0	0
M00001504C:A07	10185	0	0	0	0	0	0
M00001504C:H06	6974	0	0	0	0	0	0
M00001504D:G06	6420	0	0	0	0	0	0
M00001507A:H05	39168	0	0	0	0	0	0
M00001511A:H06	39412	0	0	0	0	0	0
M00001512A:A09	39186	0	0	0	0	0	0
M00001512D:G09	3956	0	0	1	0	0	0
M00001513A:B06	4568	0	0	0	0	0	0
M00001513C:E08	14364	0	0	0	0	0	0
M00001514C:D11	40044	0	1	0	0	0	0
M00001517A:B07	4313	0	0	0	0	0	0
M00001518C:B11	8952	0	0	0	0	0	0
M00001528A:C04	7337	0	0	0	0	0	0
M00001528A:F09	18957	0	0	0	0	0	0
M00001528B:H04	8358	0	0	0	0	0	0
M00001531A:D01	38085	0	0	0	0	0	0
M00001532B:A06	3990	1	1	0	0	0	0
M00001533A:C11	2428	0	0	1	0	0	0
M00001534A:C04	16921	0	0	0	0	0	0
M00001534A:D09	5097	0	0	0	0	0	0
M00001534A:F09	5321	0	1	0	0	2	0
M00001534C:A01	4119	0	0	0	0	0	0
M00001535A:B01	7665	0	0	0	0	0	0
M00001535A:C06	20212	0	0	0	0	0	0
M00001535A:F10	39423	0	0	0	0	0	0
M00001536A:B07	2696	0	0	0	0	3	0
M00001536A:C08	39392	0	0	0	0	0	0
M00001537A:F12	39420	0	0	0	0	0	0
M00001537B:G07	3389	0	0	0	0	0	0
M00001540A:D06	8286	0	0	0	0	0	0
M00001541A:D02	3765	0	0	0	0	0	0
M00001541A:F07	22085	0	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00001541A:H03	39174	0	0	0	0	0	0
M00001542A:A09	22113	0	0	0	0	0	0
M00001542A:E06	39453	0	0	0	0	0	0
M00001544A:E03	12170	0	0	0	0	0	0
M00001544A:G02	19829	0	0	0	0	0	0
M00001544B:B07	6974	0	0	0	0	0	0
M00001545A:C03	19255	0	0	0	0	0	0
M00001545A:D08	13864	0	0	0	0	0	0
M00001546A:G11	1267	1	0	0	0	7	0
M00001548A:E10	5892	0	0	0	0	0	0
M00001548A:H09	1058	0	0	1	0	0	0
M00001549A:B02	4015	0	0	0	0	0	0
M00001549A:D08	10944	0	0	0	0	0	0
M00001549B:F06	4193	0	0	0	0	0	0
M00001549C:E06	16347	0	0	0	0	0	0
M00001550A:A03	7239	0	0	0	0	0	0
M00001550A:G01	5175	0	0	0	0	0	0
M00001551A:B10	6268	0	0	0	0	0	0
M00001551A:F05	39180	0	0	0	0	0	0
M00001551A:G06	22390	0	0	0	0	0	0
M00001551C:G09	3266	0	0	1	0	0	0
M00001552A:B12	307	0	0	0	0	3	0
M00001552A:D11	39458	0	0	0	0	0	0
M00001552B:D04	5708	0	1	0	0	0	0
M00001553A:H06	8298	0	0	0	0	0	0
M00001553B:F12	4573	0	0	0	0	0	0
M00001553D:D10	22814	0	0	0	0	0	0
M00001555A:B02	39539	0	0	0	0	0	0
M00001555A:C01	39195	0	0	0	0	0	0
M00001555D:G10	4561	0	0	0	0	0	0
M00001556A:C09	9244	0	0	0	0	0	0
M00001556A:F11	1577	0	0	0	0	0	0
M00001556A:H01	15855	3	5	5	0	3	1
M00001556B:C08	4386	1	2	0	0	0	0
M00001556B:G02	11294	0	0	0	0	0	0
M00001557A:D02	7065	0	0	0	0	0	0
M00001557A:D02	7065	0	0	0	0	0	0
M00001557A:F01	9635	0	0	0	0	0	0
M00001557A:F03	39490	0	0	0	0	0	0
M00001557B:H10	5192	0	0	0	0	0	0
M00001557D:D09	8761	0	0	0	0	0	0
M00001558B:H11	7514	0	0	0	0	0	0
M00001560D:F10	6558	0	0	0	0	0	0
M00001561A:C05	39486	0	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00001563B:F06	102	22	38	65	7	43	10
M00001564A:B12	5053	0	0	1	0	0	0
M00001571C:H06	5749	0	0	0	0	0	0
M00001578B:E04	23001	0	0	0	0	0	0
M00001579D:C03	6539	0	0	0	0	0	0
M00001583D:A10	6293	0	0	0	0	0	0
M00001586C:C05	4623	0	0	0	0	1	0
M00001587A:B11	39380	0	0	0	0	0	0
M00001594B:H04	260	0	0	0	0	1	0
M00001597C:H02	4837	0	0	0	0	0	0
M00001597D:C05	10470	0	0	0	0	0	0
M00001598A:G03	16999	1	1	1	0	0	0
M00001601A:D08	22794	0	0	0	0	0	0
M00001604A:B10	1399	0	0	0	0	0	0
M00001604A:F05	39391	0	0	0	0	0	0
M00001607A:E11	11465	0	0	0	0	0	0
M00001608A:B03	7802	0	0	0	0	0	0
M00001608B:E03	22155	0	0	0	0	0	0
M00001614C:F10	13157	0	0	0	0	0	0
M00001617C:E02	17004	0	0	0	0	1	0
M00001619C:F12	40314	0	0	0	0	0	0
M00001621C:C08	40044	0	1	0	0	0	0
M00001623D:F10	13913	0	0	0	0	0	0
M00001624A:B06	3277	0	0	0	0	0	0
M00001624C:F01	4309	0	0	0	0	0	0
M00001630B:H09	5214	1	0	0	1	1	0
M00001644C:B07	39171	0	0	0	0	0	0
M00001645A:C12	19267	0	0	0	0	1	0
M00001648C:A01	4665	0	0	0	0	0	0
M00001657D:C03	23201	0	0	0	0	0	0
M00001657D:F08	76760	0	0	0	0	0	0
M00001662C:A09	23218	0	0	0	0	0	0
M00001663A:E04	35702	0	0	0	0	0	0
M00001669B:F02	6468	0	0	0	0	0	0
M00001670C:H02	14367	0	0	0	0	0	0
M00001673C:H02	7015	0	0	0	0	0	0
M00001675A:C09	8773	0	0	0	0	0	0
M00001676B:F05	11460	0	0	0	0	0	0
M00001677C:E10	14627	0	1	0	0	0	0
M00001677D:A07	7570	0	0	0	0	0	0
M00001678D:F12	4416	0	0	0	0	0	0
M00001679A:A06	6660	0	0	0	0	0	0
M00001679A:F10	26875	0	0	0	0	0	0
M00001679B:F01	6298	0	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00001679C:F01	78091	0	0	0	0	0	0
M00001679D:D03	10751	0	0	0	0	0	0
M00001679D:D03	10751	0	0	0	0	0	0
M00001680D:F08	10539	0	0	0	0	0	0
M00001682C:B12	17055	0	0	0	0	0	0
M00001686A:E06	4622	0	0	0	0	0	0
M00001688C:F09	5382	0	0	0	0	0	0
M00001693C:G01	4393	0	0	0	0	0	0
M00001716D:H05	67252	0	0	0	0	0	0
M00003741D:C09	40108	0	0	0	0	0	0
M00003747D:C05	11476	0	0	0	0	0	0
M00003759B:B09	697	0	0	0	0	1	0
M00003762C:B08	17076	0	0	0	0	0	0
M00003763A:F06	3108	0	0	0	0	0	0
M00003774C:A03	67907	0	0	0	0	0	0
M00003796C:D05	5619	0	0	0	0	0	0
M00003826B:A06	11350	0	0	0	0	0	0
M00003833A:E05	21877	0	0	0	0	0	0
M00003837D:A01	7899	0	0	0	0	0	0
M00003839A:D08	7798	0	0	0	0	0	0
M00003844C:B11	6539	0	0	0	0	0	0
M00003846B:D06	6874	0	0	1	0	0	0
M00003851B:D10	13595	0	0	0	0	0	0
M00003853A:D04	5619	0	0	0	0	0	0
M00003853A:F12	10515	0	0	0	0	0	0
M00003856B:C02	4622	0	0	0	0	0	0
M00003857A:G10	3389	0	0	0	0	0	0
M00003857A:H03	4718	0	0	0	0	0	0
M00003871C:E02	4573	0	0	0	0	0	0
M00003875B:F04	12977	0	0	0	0	0	0
M00003875B:F04	12977	0	0	0	0	0	0
M00003875C:G07	8479	0	0	0	0	0	1
M00003876D:E12	7798	0	0	0	0	0	0
M00003879B:C11	5345	0	0	0	2	0	1
M00003879B:D10	31587	0	0	0	0	0	0
M00003879D:A02	14507	0	0	0	0	0	0
M00003885C:A02	13576	0	0	0	0	0	0
M00003885C:A02	13576	0	0	0	0	0	0
M00003906C:E10	9285	0	0	0	0	0	0
M00003907D:A09	39809	0	0	0	0	0	0
M00003907D:H04	16317	0	0	0	0	0	0
M00003909D:C03	8672	0	0	0	0	0	0
M00003912B:D01	12532	0	0	0	0	0	0
M00003914C:F05	3900	0	0	0	0	1	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00003922A:E06	23255	0	0	0	0	0	0
M00003958A:H02	18957	0	0	0	0	0	0
M00003958A:H02	18957	0	0	0	0	0	0
M00003958C:G10	40455	0	0	0	0	0	0
M00003958C:G10	40455	0	0	0	0	0	0
M00003968B:F06	24488	0	0	0	0	0	0
M00003970C:B09	40122	0	0	0	0	0	0
M00003974D:E07	23210	0	0	0	0	0	0
M00003974D:H02	23358	0	0	0	0	0	0
M00003975A:G11	12439	0	0	0	0	0	0
M00003978B:G05	5693	0	0	0	0	0	0
M00003981A:E10	3430	0	0	0	0	1	0
M00003982C:C02	2433	0	0	0	0	0	0
M00003983A:A05	9105	0	0	0	0	0	0
M00004028D:A06	6124	0	0	0	0	0	0
M00004028D:C05	40073	0	0	0	0	0	0
M00004031A:A12	9061	0	0	0	0	0	0
M00004031A:A12	9061	0	0	0	0	0	0
M00004035C:A07	37285	0	0	0	0	0	0
M00004035D:B06	17036	0	0	0	0	0	0
M00004059A:D06	5417	0	0	0	0	0	0
M00004068B:A01	3706	0	0	0	0	0	0
M00004072B:B05	17036	0	0	0	0	0	0
M00004081C:D10	15069	0	0	0	0	0	0
M00004081C:D12	14391	0	0	0	0	0	0
M00004086D:G06	9285	0	0	0	0	0	0
M00004087D:A01	6880	0	0	0	0	0	0
M00004093D:B12	5325	1	1	0	1	0	1
M00004093D:B12	5325	1	1	0	1	0	1
M00004105C:A04	7221	0	0	0	0	0	0
M00004108A:E06	4937	0	0	0	0	0	0
M00004111D:A08	6874	0	0	1	0	0	0
M00004114C:F11	13183	0	0	0	0	0	0
M00004138B:H02	13272	0	0	0	0	0	0
M00004146C:C11	5257	0	1	0	0	0	0
M00004151D:B08	16977	0	0	0	0	0	0
M00004157C:A09	6455	0	0	0	0	0	0
M00004169C:C12	5319	0	0	0	0	0	0
M00004171D:B03	4908	0	0	0	0	0	0
M00004172C:D08	11494	0	0	0	0	0	0
M00004183C:D07	16392	0	0	0	0	0	0
M00004185C:C03	11443	0	0	0	0	0	0
M00004197D:H01	8210	0	0	0	0	0	0
M00004203B:C12	14311	0	0	0	0	0	0

Clone Name	Cluster ID	Clones in Lib15	Clones in Lib16b	Clones in Lib17	Clones in Lib18	Clones in Lib19	Clones in Lib20
M00004212B:C07	2379	0	0	0	0	0	0
M00004214C:H05	11451	0	0	0	0	0	0
M00004223A:G10	16918	0	0	0	0	0	0
M00004223B:D09	7899	0	0	0	0	0	0
M00004223D:E04	12971	0	0	0	0	0	0
M00004229B:F08	6455	0	0	0	0	0	0
M00004230B:C07	7212	0	0	0	0	0	0
M00004269D:D06	4905	0	0	0	0	0	0
M00004275C:C11	16914	0	0	0	0	0	0
M00004283B:A04	14286	0	0	0	0	0	0
M00004285B:E08	56020	0	0	0	0	0	0
M00004295D:F12	16921	0	0	0	0	0	0
M00004296C:H07	13046	0	0	0	0	0	0
M00004307C:A06	9457	0	0	0	0	0	0
M00004312A:G03	26295	0	0	0	0	0	0
M00004318C:D10	21847	0	0	0	0	0	0
M00004372A:A03	2030	0	0	0	0	0	0
M00004377C:F05	2102	0	0	0	0	0	0

Table 7

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00001340B:A06	17062	0	0	0
M00001340D:F10	11589	0	0	0
M00001341A:E12	4443	4	2	0
M00001342B:E06	39805	0	0	0
M00001343C:F10	2790	0	0	0
M00001343D:H07	23255	0	0	0
M00001345A:E01	6420	0	0	0
M00001346A:F09	5007	0	0	0
M00001346D:E03	6806	0	1	1
M00001346D:G06	5779	0	0	0
M00001346D:G06	5779	0	0	0
M00001347A:B10	13576	0	0	0
M00001348B:B04	16927	0	0	0
M00001348B:G06	16985	0	0	0
M00001349B:B08	3584	0	0	0
M00001350A:H01	7187	0	0	0
M00001351B:A08	3162	0	0	1
M00001351B:A08	3162	0	0	1
M00001352A:E02	16245	0	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00001353A:G12	8078	0	0	0
M00001353D:D10	14929	0	1	0
M00001355B:G10	14391	0	0	0
M00001357D:D11	4059	0	0	0
M00001361A:A05	4141	1	2	1
M00001361D:F08	2379	0	0	0
M00001362B:D10	5622	0	2	1
M00001362C:H11	945	0	0	0
M00001365C:C10	40132	0	0	0
M00001370A:C09	6867	0	0	0
M00001371C:E09	7172	0	0	1
M00001376B:G06	17732	2	0	0
M00001378B:B02	39833	0	0	0
M00001379A:A05	1334	0	0	0
M00001380D:B09	39886	0	0	0
M00001382C:A02	22979	1	0	0
M00001383A:C03	39648	0	0	0
M00001383A:C03	39648	0	0	0
M00001386C:B12	5178	0	0	0
M00001387A:C05	2464	0	0	0
M00001387B:G03	7587	0	0	0
M00001388D:G05	5832	0	0	0
M00001389A:C08	16269	2	0	0
M00001394A:F01	6583	0	0	0
M00001395A:C03	4016	0	0	0
M00001396A:C03	4009	2	0	0
M00001402A:E08	39563	0	0	0
M00001407B:D11	5556	0	0	0
M00001409C:D12	9577	0	0	0
M00001410A:D07	7005	0	0	0
M00001412B:B10	8551	0	0	0
M00001415A:H06	13538	0	0	0
M00001416A:H01	7674	0	0	0
M00001416B:H11	8847	1	0	0
M00001417A:E02	36393	0	0	0
M00001418B:F03	9952	0	0	0
M00001418D:B06	8526	0	0	0
M00001421C:F01	9577	0	0	0
M00001423B:E07	15066	0	0	0
M00001424B:G09	10470	0	0	0
M00001425B:H08	22195	0	0	0
M00001426D:C08	4261	0	0	0
M00001428A:H10	84182	0	0	0
M00001429A:H04	2797	0	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00001429B:A11	4635	0	0	0
M00001429D:D07	40392	0	0	0
M00001439C:F08	40054	0	0	0
M00001442C:D07	16731	0	0	0
M00001445A:F05	13532	0	0	0
M00001446A:F05	7801	0	1	0
M00001447A:G03	10717	0	0	0
M00001448D:C09	8	7	6	9
M00001448D:H01	36313	1	0	0
M00001449A:A12	5857	0	0	0
M00001449A:B12	41633	0	0	0
M00001449A:D12	3681	1	0	0
M00001449A:G10	36535	0	0	0
M00001449C:D06	86110	0	0	0
M00001450A:A02	39304	0	1	0
M00001450A:A11	32663	0	0	0
M00001450A:B12	82498	0	0	0
M00001450A:D08	27250	0	0	0
M00001452A:B04	84328	0	0	0
M00001452A:B12	86859	0	0	0
M00001452A:D08	1120	0	0	0
M00001452A:F05	85064	0	0	0
M00001452C:B06	16970	1	0	0
M00001453A:E11	16130	0	0	0
M00001453C:F06	16653	0	0	0
M00001454A:A09	83103	0	0	0
M00001454B:C12	7005	0	0	0
M00001454D:G03	689	0	0	1
M00001455A:E09	13238	0	0	0
M00001455B:E12	13072	0	0	0
M00001455D:F09	9283	0	0	0
M00001455D:F09	9283	0	0	0
M00001460A:F06	2448	0	0	0
M00001460A:F12	39498	0	0	0
M00001461A:D06	1531	0	0	1
M00001463C:B11	19	17	32	31
M00001465A:B11	10145	0	0	0
M00001466A:E07	4275	0	0	0
M00001467A:B07	38759	0	0	0
M00001467A:D04	39508	0	0	0
M00001467A:D08	16283	0	0	0
M00001467A:D08	16283	0	0	0
M00001467A:E10	39442	0	0	0
M00001468A:F05	7589	0	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00001469A:C10	12081	0	0	0
M00001469A:H12	19105	0	0	0
M00001470A:B10	1037	0	0	0
M00001470A:C04	39425	0	0	0
M00001471A:B01	39478	0	0	0
M00001481D:A05	7985	0	0	0
M00001490B:C04	18699	0	0	0
M00001494D:F06	7206	0	0	0
M00001497A:G02	2623	1	0	0
M00001499B:A11	10539	0	1	0
M00001500A:C05	5336	0	0	0
M00001500A:E11	2623	1	0	0
M00001500C:E04	9443	0	0	0
M00001501D:C02	9685	0	0	0
M00001504C:A07	10185	0	0	0
M00001504C:H06	6974	0	0	0
M00001504D:G06	6420	0	0	0
M00001507A:H05	39168	0	0	0
M00001511A:H06	39412	0	0	0
M00001512A:A09	39186	0	0	0
M00001512D:G09	3956	0	0	0
M00001513A:B06	4568	0	0	0
M00001513C:E08	14364	0	0	0
M00001514C:D11	40044	0	0	0
M00001517A:B07	4313	0	0	0
M00001518C:B11	8952	0	0	0
M00001528A:C04	7337	1	2	2
M00001528A:F09	18957	0	0	0
M00001528B:H04	8358	0	0	0
M00001531A:D01	38085	0	0	0
M00001532B:A06	3990	0	0	0
M00001533A:C11	2428	0	0	0
M00001534A:C04	16921	0	0	0
M00001534A:D09	5097	0	0	0
M00001534A:F09	5321	4	7	6
M00001534C:A01	4119	0	0	0
M00001535A:B01	7665	0	2	4
M00001535A:C06	20212	0	0	0
M00001535A:F10	39423	0	0	0
M00001536A:B07	2696	0	0	0
M00001536A:C08	39392	0	0	0
M00001537A:F12	39420	0	0	0
M00001537B:G07	3389	0	0	0
M00001540A:D06	8286	0	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00001541A:D02	3765	0	0	0
M00001541A:F07	22085	0	0	0
M00001541A:H03	39174	0	0	0
M00001542A:A09	22113	0	0	0
M00001542A:E06	39453	0	0	0
M00001544A:E03	12170	0	0	0
M00001544A:G02	19829	0	0	0
M00001544B:B07	6974	0	0	0
M00001545A:C03	19255	0	0	0
M00001545A:D08	13864	0	0	0
M00001546A:G11	1267	0	0	0
M00001548A:E10	5892	0	1	0
M00001548A:H09	1058	1	3	0
M00001549A:B02	4015	0	1	0
M00001549A:D08	10944	1	0	0
M00001549B:F06	4193	0	0	0
M00001549C:E06	16347	0	0	0
M00001550A:A03	7239	0	1	0
M00001550A:G01	5175	1	0	0
M00001551A:B10	6268	0	0	1
M00001551A:F05	39180	0	0	0
M00001551A:G06	22390	0	0	1
M00001551C:G09	3266	0	0	0
M00001552A:B12	307	6	11	4
M00001552A:D11	39458	0	0	0
M00001552B:D04	5708	0	0	0
M00001553A:H06	8298	0	0	0
M00001553B:F12	4573	0	0	0
M00001553D:D10	22814	0	0	0
M00001555A:B02	39539	0	0	0
M00001555A:C01	39195	0	0	0
M00001555D:G10	4561	0	0	0
M00001556A:C09	9244	0	1	0
M00001556A:F11	1577	0	0	2
M00001556A:H01	15855	1	1	0
M00001556B:C08	4386	3	0	1
M00001556B:G02	11294	0	0	0
M00001557A:D02	7065	0	0	0
M00001557A:D02	7065	0	0	0
M00001557A:F01	9635	0	0	0
M00001557A:F03	39490	0	0	0
M00001557B:H10	5192	0	0	0
M00001557D:D09	8761	0	0	0
M00001558B:H11	7514	0	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00001560D:F10	6558	0	0	0
M00001561A:C05	39486	0	0	0
M00001563B:F06	102	2	1	2
M00001564A:B12	5053	0	0	0
M00001571C:H06	5749	0	0	0
M00001578B:E04	23001	0	0	0
M00001579D:C03	6539	0	0	0
M00001583D:A10	6293	0	0	0
M00001586C:C05	4623	0	0	0
M00001587A:B11	39380	0	0	0
M00001594B:H04	260	1	0	0
M00001597C:H02	4837	1	0	0
M00001597D:C05	10470	0	0	0
M00001598A:G03	16999	4	2	6
M00001601A:D08	22794	0	0	0
M00001604A:B10	1399	6	3	3
M00001604A:F05	39391	0	0	0
M00001607A:E11	11465	0	0	0
M00001608A:B03	7802	0	0	0
M00001608B:E03	22155	0	0	0
M00001614C:F10	13157	0	0	0
M00001617C:E02	17004	0	0	0
M00001619C:F12	40314	0	0	0
M00001621C:C08	40044	0	0	0
M00001623D:F10	13913	0	0	0
M00001624A:B06	3277	0	0	0
M00001624C:F01	4309	0	0	0
M00001630B:H09	5214	0	1	2
M00001644C:B07	39171	0	0	0
M00001645A:C12	19267	0	0	0
M00001648C:A01	4665	0	0	0
M00001657D:C03	23201	0	0	0
M00001657D:F08	76760	0	0	0
M00001662C:A09	23218	0	0	0
M00001663A:E04	35702	0	0	0
M00001669B:F02	6468	0	0	0
M00001670C:H02	14367	0	0	0
M00001673C:H02	7015	0	0	0
M00001675A:C09	8773	0	0	0
M00001676B:F05	11460	2	0	0
M00001677C:E10	14627	0	0	0
M00001677D:A07	7570	0	0	0
M00001678D:F12	4416	1	2	0
M00001679A:A06	6660	0	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00001679A:F10	26875	0	0	0
M00001679B:F01	6298	0	0	0
M00001679C:F01	78091	0	0	0
M00001679D:D03	10751	0	0	0
M00001679D:D03	10751	0	0	0
M00001680D:F08	10539	0	1	0
M00001682C:B12	17055	0	0	0
M00001686A:E06	4622	0	0	0
M00001688C:F09	5382	0	0	0
M00001693C:G01	4393	0	0	0
M00001716D:H05	67252	0	0	0
M00003741D:C09	40108	0	0	0
M00003747D:C05	11476	0	0	0
M00003759B:B09	697	0	0	0
M00003762C:B08	17076	0	0	0
M00003763A:F06	3108	0	0	0
M00003774C:A03	67907	0	0	0
M00003796C:D05	5619	0	1	0
M00003826B:A06	11350	0	0	0
M00003833A:E05	21877	0	0	0
M00003837D:A01	7899	0	0	0
M00003839A:D08	7798	0	0	0
M00003844C:B11	6539	0	0	0
M00003846B:D06	6874	0	0	0
M00003851B:D10	13595	0	0	0
M00003853A:D04	5619	0	1	0
M00003853A:F12	10515	0	0	1
M00003856B:C02	4622	0	0	0
M00003857A:G10	3389	0	0	0
M00003857A:H03	4718	0	0	0
M00003871C:E02	4573	0	0	0
M00003875B:F04	12977	0	0	0
M00003875B:F04	12977	0	0	0
M00003875C:G07	8479	1	0	0
M00003876D:E12	7798	0	0	0
M00003879B:C11	5345	4	8	3
M00003879B:D10	31587	0	0	0
M00003879D:A02	14507	0	0	0
M00003885C:A02	13576	0	0	0
M00003885C:A02	13576	0	0	0
M00003906C:E10	9285	0	0	0
M00003907D:A09	39809	0	0	0
M00003907D:H04	16317	0	0	0
M00003909D:C03	8672	0	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00003912B:D01	12532	0	0	0
M00003914C:F05	3900	0	1	0
M00003922A:E06	23255	0	0	0
M00003958A:H02	18957	0	0	0
M00003958A:H02	18957	0	0	0
M00003958C:G10	40455	0	0	0
M00003958C:G10	40455	0	0	0
M00003968B:F06	24488	0	0	0
M00003970C:B09	40122	0	0	0
M00003974D:E07	23210	0	0	0
M00003974D:H02	23358	0	0	0
M00003975A:G11	12439	0	0	0
M00003978B:G05	5693	0	0	0
M00003981A:E10	3430	0	0	0
M00003982C:C02	2433	2	4	0
M00003983A:A05	9105	0	0	0
M00004028D:A06	6124	0	0	0
M00004028D:C05	40073	0	1	0
M00004031A:A12	9061	0	0	0
M00004031A:A12	9061	0	0	0
M00004035C:A07	37285	0	0	0
M00004035D:B06	17036	0	0	0
M00004059A:D06	5417	0	0	0
M00004068B:A01	3706	0	0	0
M00004072B:B05	17036	0	0	0
M00004081C:D10	15069	0	0	0
M00004081C:D12	14391	0	0	0
M00004086D:G06	9285	0	0	0
M00004087D:A01	6880	0	0	0
M00004093D:B12	5325	0	0	0
M00004093D:B12	5325	0	0	0
M00004105C:A04	7221	0	0	0
M00004108A:E06	4937	0	0	0
M00004111D:A08	6874	0	0	0
M00004114C:F11	13183	0	0	0
M00004138B:H02	13272	0	0	0
M00004146C:C11	5257	0	0	1
M00004151D:B08	16977	0	0	0
M00004157C:A09	6455	0	0	0
M00004169C:C12	5319	0	0	0
M00004171D:B03	4908	0	0	0
M00004172C:D08	11494	0	0	0
M00004183C:D07	16392	0	0	0
M00004185C:C03	11443	2	0	0

Clone Name	Cluster ID	Clones in Lib12	Clones in Lib13	Clones in Lib14
M00004197D:H01	8210	0	0	0
M00004203B:C12	14311	0	0	0
M00004212B:C07	2379	0	0	0
M00004214C:H05	11451	0	0	0
M00004223A:G10	16918	0	0	0
M00004223B:D09	7899	0	0	0
M00004223D:E04	12971	0	0	0
M00004229B:F08	6455	0	0	0
M00004230B:C07	7212	0	0	1
M00004269D:D06	4905	0	0	0
M00004275C:C11	16914	0	0	0
M00004283B:A04	14286	0	0	0
M00004285B:E08	56020	0	0	0
M00004295D:F12	16921	0	0	0
M00004296C:H07	13046	0	0	0
M00004307C:A06	9457	1	0	0
M00004312A:G03	26295	0	0	0
M00004318C:D10	21847	0	0	0
M00004372A:A03	2030	0	0	0
M00004377C:F05	2102	0	0	0

Example 5: Polynucleotides Differentially Expressed in High Metastatic Potential Breast Cancer Cells Versus Low Metastatic Breast Cancer Cells

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential breast cancer tissue and low metastatic breast cancer cells. Expression of these sequences in breast cancer can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit

metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following table summarizes identified polynucleotides with differential expression between high metastatic potential breast cancer cells and low metastatic potential breast cancer cells.

Table 8. Differentially expressed polynucleotides: High metastatic potential breast cancer vs. low metastatic breast cancer cells

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
9	High Breast > Low Breast (Lib3 > Lib4)	2623	31	4	7.561356
42	High Breast > Low Breast (Lib3 > Lib4)	307	196	75	2.549721
52	High Breast > Low Breast (Lib3 > Lib4)	19	1364	525	2.534854
62	High Breast > Low Breast (Lib3 > Lib4)	2623	31	4	7.561356
65	High Breast > Low Breast (Lib3 > Lib4)	5749	9	0	8.780930
66	High Breast > Low Breast (Lib3 > Lib4)	6455	6	0	5.853953
68	High Breast > Low Breast (Lib3 > Lib4)	6455	6	0	5.853953
114	High Breast > Low Breast (Lib3 > Lib4)	2030	32	4	7.805271
123	High Breast > Low Breast (Lib3 > Lib4)	3389	13	2	6.341782
144	High Breast > Low Breast (Lib3 > Lib4)	4623	12	2	5.853953
172	High Breast > Low Breast (Lib3 > Lib4)	102	278	116	2.338217
178	High Breast > Low Breast (Lib3 > Lib4)	3681	10	1	9.756589
214	High Breast > Low Breast (Lib3 > Lib4)	3900	8	1	7.805271
219	High Breast > Low Breast (Lib3 > Lib4)	3389	13	2	6.341782
223	High Breast > Low Breast (Lib3 > Lib4)	1399	19	7	2.648217
258	High Breast > Low Breast (Lib3 > Lib4)	4837	10	0	9.756589
317	High Breast > Low Breast (Lib3 > Lib4)	1577	25	3	8.130490
379	High Breast > Low Breast (Lib3 > Lib4)	260	27	2	13.17139
4	Low Breast > High Breast (Lib4 > Lib3)	3706	22	4	5.637215
39	Low Breast > High Breast (Lib4 > Lib3)	4016	6	0	6.149690
74	Low Breast > High Breast (Lib4 > Lib3)	6268	18	3	6.149690
81	Low Breast > High Breast (Lib4 > Lib3)	40392	8	1	8.199586
130	Low Breast > High Breast (Lib4 > Lib3)	13183	7	0	7.174638
157	Low Breast > High Breast (Lib4 > Lib3)	5417	9	0	9.224535

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
162	Low Breast > High Breast (Lib4 > Lib3)	9685	7	0	7.174638
183	Low Breast > High Breast (Lib4 > Lib3)	7337	16	3	5.466391
202	Low Breast > High Breast (Lib4 > Lib3)	6124	9	1	9.224535
298	Low Breast > High Breast (Lib4 > Lib3)	1037	22	4	5.637215
338	Low Breast > High Breast (Lib4 > Lib3)	689	36	17	2.170478
384	Low Breast > High Breast (Lib4 > Lib3)	697	72	30	2.459876
386	Low Breast > High Breast (Lib4 > Lib3)	4568	9	0	9.224535
388	Low Breast > High Breast (Lib4 > Lib3)	5622	13	2	6.662164

Example 6: Polynucleotides Differentially Expressed in High Metastatic Potential Lung Cancer Cells Versus Low Metastatic Lung Cancer Cells

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential lung cancer tissue and low metastatic lung cancer cells. Expression of these sequences in lung cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells are associated can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following table summarizes identified polynucleotides with differential expression between high metastatic potential lung cancer cells and low metastatic potential lung cancer cells:

Table 9 Differentially expressed polynucleotides: High metastatic potential lung cancer vs. low metastatic lung cancer cells

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
400	High Lung > Low Lung (Lib8 > Lib 9)	14929	23	16	2.008868
9	High Lung > Low Lung (Lib8 > Lib9)	2623	6	1	8.384840
34	High Lung > Low Lung (Lib8 > Lib9)	5832	5	0	6.987366
42	High Lung > Low Lung (Lib8 > Lib9)	307	79	27	4.088903
62	High Lung > Low Lung (Lib8 > Lib9)	2623	6	1	8.384840
74	High Lung > Low Lung (Lib8 > Lib9)	6268	5	0	6.987366
106	High Lung > Low Lung (Lib8 > Lib9)	10717	8	0	11.17978
119	High Lung > Low Lung (Lib8 > Lib9)	8	1355	122	15.52111
361	High Lung > Low Lung (Lib8 > Lib9)	1120	5	0	6.987366
369	High Lung > Low Lung (Lib8 > Lib9)	2790	6	0	8.384840
371	High Lung > Low Lung (Lib8 > Lib9)	8847	6	1	8.384840
379	High Lung > Low Lung (Lib8 > Lib9)	260	15	0	20.96210
395	High Lung > Low Lung (Lib8 > Lib9)	13538	9	1	12.57726
135	Low Lung > High Lung (Lib9 > Lib8)	36313	30	1	21.46731
154	Low Lung > High Lung (Lib9 > Lib8)	5345	27	6	3.220097
160	Low Lung > High Lung (Lib9 > Lib8)	4386	21	3	5.009039
260	Low Lung > High Lung (Lib9 > Lib8)	4141	27	4	4.830145
308	Low Lung > High Lung (Lib9 > Lib8)	15855	213	12	12.70149
323	Low Lung > High Lung (Lib9 > Lib8)	5257	25	5	3.577885
349	Low Lung > High Lung (Lib9 > Lib8)	2797	14	1	10.01807
381	Low Lung > High Lung (Lib9 > Lib8)	2428	19	2	6.797982

Example 7: Polynucleotides Differentially Expressed in High Metastatic Potential Colon Cancer Cells Versus Low Metastatic Colon Cancer Cells

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential colon cancer tissue and low metastatic colon cancer cells. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher

expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following table summarizes identified polynucleotides with differential expression between high metastatic potential colon cancer cells and low metastatic potential colon cancer cells:

Table 10 Differentially expressed polynucleotides: High metastatic potential colon cancer vs. low metastatic colon cancer cells

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
1	High Colon > Low Colon (Lib1 > Lib2)	6660	7	0	6.489973
176	High Colon > Low Colon (Lib1 > Lib2)	3765	19	6	2.935940
241	High Colon > Low Colon (Lib1 > Lib2)	4275	11	2	5.099264
362	High Colon > Low Colon (Lib1 > Lib2)	6420	8	0	7.417112
374	High Colon > Low Colon (Lib1 > Lib2)	6420	8	0	7.417112
39	Low Colon > High Colon (Lib2 > Lib1)	4016	14	5	3.020043
97	Low Colon > High Colon (Lib2 > Lib1)	945	21	9	2.516702
134	Low Colon > High Colon (Lib2 > Lib1)	2464	19	5	4.098630
317	Low Colon > High Colon (Lib2 > Lib1)	1577	40	12	3.595289
357	Low Colon > High Colon (Lib2 > Lib1)	4309	13	4	3.505407

Example 8: Polynucleotides Differentially Expressed at Higher Levels in High Metastatic Potential Colon Cancer Patient Tissue Versus Normal Patient Tissue

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells are associated can be indicative of increased expression of genes or regulatory sequences involved in the advanced disease state which involves processes such as angiogenesis, dedifferentiation, cell replication, and metastasis. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following table summarizes identified polynucleotides with differential expression between high metastatic potential colon cancer cells and normal colon cells:

Table 11: Differentially expressed polynucleotides: High metastatic potential colon tissue vs. normal colon tissue

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
52	High Colon Metastasis Tissue > Normal Colon Tissue of UC#3 (Lib20 > Lib18)	19	10	0	11.69918
52	High Colon Metastasis Tissue > Normal Tissue in UC#2 (Lib17 > Lib15)	19	13	2	6.025646
172	High Colon Metastasis Tissue > Normal Tissue in UC#2 (Lib17 > Lib15)	102	65	22	2.738930

Example 9: Polynucleotides Differentially Expressed at Higher Levels in High Colon Tumor Potential Patient Tissue Versus Metastasized Colon Cancer Patient Tissue

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high tumor potential colon cancer tissue and cells derived from high metastatic potential colon cancer cells. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the transformation of precancerous tissue to malignant tissue. This information can be useful in the prevention of achieving the advanced malignant state in these tissues, and can be important in risk assessment for a patient.

The following table summarizes identified polynucleotides with differential expression between high tumor potential colon cancer tissue and cells derived from high metastatic potential colon cancer cells:

Table 12: Differentially expressed polynucleotides: High tumor potential colon tissue vs. metastatic colon tissue

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
52	High Colon Tumor Tissue > Metastasis Tissue of UC#3 (Lib19 > Lib20)	19	69	10	5.160829
119	High Colon Tumor Tissue > Metastasis Tissue of UC#3 (Lib19 > Lib20)	8	14	1	10.47124

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
172	High Colon Tumor Tissue > Metastasis Tissue of UC#3 (Lib19 > Lib20)	102	43	10	3.216168

Example 10: Polynucleotides Differentially Expressed at Higher Levels in High Tumor Potential Colon Cancer Patient Tissue Versus Normal Patient Tissue

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high tumor potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. For example, sequences that are highly expressed in the potential colon cancer cells are associated with or can be indicative of increased expression of genes or regulatory sequences involved in early tumor progression. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The following table summarizes identified polynucleotides with differential expression between high metastatic potential colon cancer cells and normal colon cells:

Table 13: Differentially expressed polynucleotides: High tumor potential colon tissue vs. normal colon tissue

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
52	High Colon Tumor Tissue > Normal Tissue of UC#2 (Lib16 > Lib15)	19	13	2	6.255508
288	High Colon Tumor Tissue > Normal Tissue of UC#2 (Lib16 > Lib15)	1267	7	0	6.125253
52	High Colon Tumor Tissue > Normal Tissue of UC#3 (Lib19 > Lib18)	19	69	0	60.37750

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
119	High Colon Tumor Tissue > Normal Tissue of UC#3 (Lib19 > Lib18)	8	14	1	12.2505 0
172	High Colon Tumor Tissue > Normal Tissue of UC#3 (Lib19 > Lib18)	102	43	7	5.37522 2

Example 11: Polynucleotides Differentially Expressed Across Multiple Libraries

A number of polynucleotide sequences have been identified that are differentially expressed between cancerous cells and normal cells across all three tissue types tested (*i.e.*, breast, colon, and lung). Expression of these sequences in a tissue or any origin can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. These polynucleotides can also serve as non-tissue specific markers of, for example, risk of metastasis of a tumor. The following table summarizes identified polynucleotides that were differentially expressed but without tissue type-specificity in the breast, colon, and lung libraries tested.

Table 14: Polynucleotides Differentially Expressed Across Multiple Library Comparisons

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
9	High Breast > Low Breast (Lib3 > Lib4)	2623	31	4	7.561356
	High Lung > Low Lung (Lib8 > Lib9)	2623	6	1	8.384840
39	Low Breast > High Breast (Lib4 > Lib3)	4016	6	0	6.149690
	Low Colon > High Colon (Lib2 > Lib1)	4016	14	5	3.020043
42	High Breast > Low Breast (Lib3 > Lib4)	307	196	75	2.549721
	High Lung > Low Lung (Lib8 > Lib9)	307	79	27	4.088903
52	High Breast > Low Breast (Lib3 > Lib4)	19	1364	525	2.534854
	High Colon Metastasis Tissue > Normal Colon Tissue of UC#3 (Lib20 > Lib18)	19	10	0	11.69918
	High Colon Metastasis Tissue > Normal Tissue in UC#2 (Lib17 > Lib15)	19	13	2	6.025646
	High Colon Tumor Tissue > Metastasis Tissue of UC#3 (Lib19 > Lib20)	19	69	10	5.160829
	High Colon Tumor Tissue > Normal	19	13	2	6.255508

SEQ ID NO.	Differential Expression	Cluster ID	Clones in 1st Library	Clones in 2nd Library	Ratio
	Tissue of UC#2 (Lib16 > Lib15)				
	High Colon Tumor Tissue > Normal Tissue of UC#3 (Lib19 > Lib18)	19	69	0	60.37750
62	High Breast > Low Breast (Lib3 > Lib4)	2623	31	4	7.561356
	High Lung > Low Lung (Lib8 > Lib9)	2623	6	1	8.384840
74	High Lung > Low Lung (Lib8 > Lib9)	6268	5	0	6.987366
	Low Breast > High Breast (Lib4 > Lib3)	6268	18	3	6.149690
119	High Colon Tumor Tissue > Metastasis Tissue of UC#3 (Lib19 > Lib20)	8	14	1	10.47124
	High Colon Tumor Tissue > Normal Tissue of UC#3 (Lib19 > Lib18)	8	14	1	12.25050
	High Lung > Low Lung (Lib8 > Lib9)	8	1355	122	15.52111
172	High Breast > Low Breast (Lib3 > Lib4)	102	278	116	2.338217
	High Colon Metastasis Tissue > Normal Tissue in UC#2 (Lib17 > Lib15)	102	65	22	2.738930
	High Colon Tumor Tissue > Metastasis Tissue of UC#3 (Lib19 > Lib20)	102	43	10	3.216168
	High Colon Tumor Tissue > Normal Tissue of UC#3 (Lib19 > Lib18)	102	43	7	5.375222
317	High Breast > Low Breast (Lib3 > Lib4)	1577	25	3	8.130490
	Low Colon > High Colon (Lib2 > Lib1)	1577	40	12	3.595289
379	High Breast > Low Breast (Lib3 > Lib4)	260	27	2	13.17139
	High Lung > Low Lung (Lib8 > Lib9)	260	15	0	20.96210

Example 12: Polynucleotides Exhibiting Colon-Specific Expression

The cDNA libraries described herein were also analyzed to identify those polynucleotides that were specifically expressed in colon cells or tissue, *i.e.*, the polynucleotides were identified in libraries prepared from colon cell lines or tissue, but not in libraries of breast or lung origin. The polynucleotides that were expressed in a colon cell line and/or in colon tissue, but were present in the breast or lung cDNA libraries described herein, are shown in Table 15.

Table 15 Polynucleotides specifically expressed in colon cells.

SEQ ID NO.	Cluster	Clones in 1st Library	Clones in 2nd Library	SEQ ID NO.	Cluster	Clones in 1st Library	Clones in 2nd Library
5	36535	2	0	229	39648	2	0
13	27250	2	0	231	85064	1	0
19	16283	3	0	234	39391	2	0
24	16918	4	0	236	39498	2	0
26	40108	2	0	242	22113	3	0
32	32663	1	1	247	19255	2	0
43	39833	2	0	252	22814	3	0
47	18957	3	0	253	39563	2	0
48	39508	2	0	254	39420	2	0
56	7005	8	2	257	39412	2	0
58	18957	3	0	261	38085	2	0
59	18957	3	0	265	40054	1	0
60	16283	3	0	266	39423	2	0
64	13238	4	1	267	39453	2	0
70	39442	2	0	270	78091	1	0
71	17036	4	0	276	39168	2	0
73	7005	8	2	277	39458	2	0
83	11476	6	0	278	14391	3	1
86	39425	2	0	279	39195	2	0
94	21847	2	1	282	12977	5	0
100	16731	3	1	284	14391	3	1
101	12439	4	0	290	16347	4	0
113	17055	4	0	293	39478	2	0
120	67907	1	0	294	39392	2	0
121	12081	4	0	297	39180	2	0
124	39174	2	0	299	6867	7	3
126	8210	2	6	301	41633	1	1
128	40455	2	0	302	23218	3	0
139	22195	3	0	303	39380	2	0
143	86859	1	0	309	84328	1	0
150	8672	4	4	314	14367	3	0
153	16977	4	0	320	39886	2	0
156	17036	4	0	324	9061	5	2
159	40044	2	0	327	16653	3	1
161	40044	2	0	328	16985	4	0
163	22155	3	0	329	12977	5	0
166	15066	4	0	330	9061	5	2
170	11465	5	0	333	16392	3	0
176	3765	19	6	342	39486	2	0

SEQ ID NO.	Cluster	Clones in 1st Library	Clones in 2nd Library	SEQ ID NO.	Cluster	Clones in 1st Library	Clones in 2nd Library
181	86110	1	0	344	6874	6	3
182	39648	2	0	345	6874	6	3
185	17076	4	0	353	11494	4	0
186	22794	2	0	354	17062	3	0
187	39171	2	0	355	16245	4	0
194	40455	2	0	356	83103	1	0
199	16317	3	0	358	13072	4	1
210	39186	2	0	366	14364	1	0
211	40122	2	0	368	84182	1	0
218	26295	2	0	372	56020	1	0
222	4665	5	9	389	7514	5	3
226	82498	1	0	391	7570	5	3
227	35702	2	0	393	23210	3	0

In addition to the above, SEQ ID NOS:159 and 161 were each present in one clone in each of Lib16 (Normal Colon Tumor Tissue), and SEQ ID NOS:344 and 345 were each present in one clone in Lib17 (High Colon Metastasis Tissue). No clones corresponding to the colon-specific polynucleotides in the table above were present in any of Libraries 3, 4, 8, or 9. The polynucleotide provided above can be used as markers of cells of colon origin, and find particular use in reference arrays, as described above.

Example 13: Identification of Contiguous Sequences Having a Polynucleotide of the Invention

The novel polynucleotides were used to screen publicly available and proprietary databases to determine if any of the polynucleotides of SEQ ID NOS:1-404 would facilitate identification of a contiguous sequence, *e.g.*, the polynucleotides would provide sequence that would result in 5' extension of another DNA sequence, resulting in production of a longer contiguous sequence composed of the provided polynucleotide and the other DNA sequence(s). Contiging was performed using the AssemblyLign program with the following parameters: 1) Overlap: Minimum Overlap Length: 30; % Stringency: 50; Minimum Repeat Length: 30; Alignment: gap creation penalty: 1.00, gap extension penalty: 1.00; 2) Consensus: % Base designation threshold: 80.

Using these parameters, 44 polynucleotides provided contiged sequences. These contiged sequences are provided as SEQ ID NOS:801-844. The contiged sequences can be correlated with the sequences of SEQ ID NOS:1-404 upon which the contiged sequences are based by identifying those sequences of SEQ ID NOS:1-404 and the contiged sequences of SEQ ID NOS:801-844 that share the same clone name in Table 1. It should be noted that of these 44 sequences that provided a contiged sequence, the following members of that group of 44 did not contig using the overlap settings indicated in parentheses (Stringency/Overlap): SEQ ID NO:804 (30%/10); SEQ ID NO:810 (20%/20); SEQ ID NO:812 (30%/10); SEQ ID NO:814 (40%/20); SEQ ID NO:816 (30%/10); SEQ ID NO:832 (30%/10); SEQ ID NO:840 (20%/20); SEQ ID NO:841 (40%/20). To generalize, the indicated polynucleotides did not contig using a minimum 20% stringency, 10 overlap. There was a corresponding increase in the number of degenerate codons in these sequences.

The contiged sequences (SEQ ID NO:801-844) thus represent longer sequences that encompass a polynucleotide sequence of the invention. The contiged sequences were then translated in all three reading frames to determine the best alignment with individual sequences using the BLAST programs as described above for SEQ ID NOS:1-404 and the validation sequences SEQ ID NOS:405-800. Again the sequences were masked using the XBLAST profram for masking low complexity as described above in Example 1 (Table 2). Several of the contiged sequences were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 16). Thus the invention encompasses fragments, fusions, and variants of such polynucleotides that retain biological activity associated with the protein family and/or functional domain identified herein.

Table 16. Profile hits using contiged sequences

SEQ ID NO.	**Sequence Name**	**Profile**	**Start (Stop)**	**Score**
809	Contig_RTA00000177AF.n.18.3. Seq_THC123051	ATPases	778 (1612)	6040
824	Contig_RTA00000187AF.g.24.1.	homeobox	531	12080

	Seq_THC168636		(707)	
824	Contig_RTA00000187AF.g.24.1. Seq_THC168636	MAP kinase kinase	769 (1494)	5784
833	Contig_RTA00000190AF.j.4.1. Seq_THC228776	protein kinase	170 (1010)	5027
833	Contig_RTA00000190AF.j.4.1. Seq_THC228776	protein kinase	170 (1010)	5027

All stop/start sequences are provided in the forward direction.

The profiles for the ATPases (AAA) and protein kinase families are described above in Example 2. The homeobox and MAP kinase kinase protein families are described further below.

Homeobox domain. The 'homeobox' is a protein domain of 60 amino acids (Gehring In: Guidebook to the Homeobox Genes, Duboule D., Ed., pp1-10, Oxford University Press, Oxford, (1994); Buerglin In: Guidebook to the Homeobox Genes, pp25-72, Oxford University Press, Oxford, (1994); Gehring *Trends Biochem. Sci.* (1992) *17*:277-280; Gehring *et al Annu. Rev. Genet.* (1986) *20*:147-173; Schofield *Trends Neurosci.* (1987) *10*:3-6; http://copan.bioz.unibas.ch/ homeo.html) first identified in number of Drosophila homeotic and segmentation proteins. It is extremely well conserved in many other animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Several proteins that contain a homeobox domain play an important role in development. Most of these proteins are sequence-specific DNA-binding transcription factors. The homeobox domain is also very similar to a region of the yeast mating type proteins. These are sequence-specific DNA-binding proteins that act as master switches in yeast differentiation by controlling gene expression in a cell type-specific fashion.

A schematic representation of the homeobox domain is shown below. The helix-turn-helix region is shown by the symbols 'H' (for helix), and 't' (for turn).

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxHHHHHHHHtttHHHHHHHHHxxxxxxxxxx
1                                                          60
```

The pattern detects homeobox sequences 24 residues long and spans positions 34 to 57 of the homeobox domain.

MAP kinase kinase (MAPKK). MAP kinases (MAPK) are involved in signal transduction, and are important in cell cycle and cell growth controls. The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast, invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates. MAPKK regulation studies have led to the discovery of at least four MAPKK convergent pathways in higher organisms. One of these is similar to the yeast pheromone response pathway which includes the ste11 protein kinase. Two other pathways require the activation of either one or both of the serine/threonine kinase-encoded oncogenes c-Raf-1 and c-Mos. Additionally, several studies suggest a possible effect of the cell cycle control regulator cyclin-dependent kinase 1 (cdc2) on MAPKK activity. Finally, MAPKKs are apparently essential transducers through which signals must pass before reaching the nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) *79*:193-207; Nishida *et al.*, *Trends Biochem Sci* (1993) *18*:128-31; Ruderman *Curr Opin Cell Biol* (1993) *5*:207-13; Dhanasekaran *et al.*, *Oncogene* (1998) *17*:1447-55; Kiefer *et al.*, *Biochem Soc Trans* (1997) *25*:491-8; and Hill, *Cell Signal* (1996) *8*:533-44.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and

modifications may be made thereto without departing from the spirit or scope of the appended claims.

Deposit Information:

The following materials were deposited with the American Type Culture Collection: CMCC = (Chiron Master Culture Collection)

Cell Lines Deposited with ATCC

Cell Line	Deposit Date	ATCC Accession No.	CMCC Accession No.
KM12L4-A	March 19, 1998	CRL-12496	11606
Km12C	May 15, 1998	CRL-12533	11611
MDA-MB-231	May 15, 1998	CRL-12532	10583
MCF-7	October 9, 1998	CRL-12584	10377

CDNA Library Deposits

cDNA Library ES1 - ATCC# 207023
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001395A:C03	4016	79.A1.sp6:130016.Seq
M00001395A:C03	4016	RTA00000118A.c.4.1
M00001449A:D12	3681	RTA00000131A.g.15.2
M00001449A:D12	3681	79.E1.sp6:130064.Seq
M00001452A:D08	1120	79.C2.sp6:130041.Seq
M00001452A:D08	1120	RTA00000118A.p.15.3
M00001513A:B06	4568	79.D4.sp6:130055.Seq
M00001513A:B06	4568	RTA00000122A.d.15.3
M00001517A:B07	4313	79.F4.sp6:130079.Seq
M00001517A:B07	4313	RTA00000122A.n.3.1
M00001533A:C11	2428	RTA00000123A.l.21.1
M00001533A:C11	2428	79.A5.sp6:130020.Seq
M00001533A:C11	2428	RTA00000123A.l.21.1.Seq_THC205063
M00001542A:A09	22113	79.F5.sp6:130080.Seq
M00001542A:A09	22113	RTA00000125A.c.7.1

cDNA Library ES2 - ATCC# 207024
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name

M00001343C:F10	2790	80.E1.sp6:130256.Seq
M00001343C:F10	2790	RTA00000177AF.e.2.1.Seq_THC229461
M00001343C:F10	2790	RTA00000177AF.e.2.1
M00001343D:H07	23255	100.C1.sp6:131446.Seq
M00001343D:H07	23255	RTA00000177AF.e.14.3.Seq_THC228776
M00001343D:H07	23255	80.F1.sp6:130268.Seq
M00001343D:H07	23255	RTA00000177AF.e.14.3
M00001345A:E01	6420	172.E1.sp6:133925.Seq
M00001345A:E01	6420	RTA00000177AF.f.10.3
M00001345A:E01	6420	RTA00000177AF.f.10.3.Seq_THC226443
M00001345A:E01	6420	80.G1.sp6:130280.Seq
M00001347A:B10	13576	80.D2.sp6:130245.Seq
M00001347A:B10	13576	100.E1.sp6:131470.Seq
M00001347A:B10	13576	RTA00000177AF.g.16.1
M00001353A:G12	8078	80.E3.sp6:130258.Seq
M00001353A:G12	8078	RTA00000177AR.l.13.1
M00001353A:G12	8078	172.C3.sp6:133903.Seq
M00001353D:D10	14929	RTA00000177AF.m.1.2
M00001353D:D10	14929	80.F3.sp6:130270.Seq
M00001353D:D10	14929	172.D3.sp6:133915.Seq
M00001361A:A05	4141	80.B4.sp6:130223.Seq
M00001361A:A05	4141	RTA00000177AF.p.20.3
M00001362B:D10	5622	80.D4.sp6:130247.Seq
M00001362B:D10	5622	RTA00000178AF.a.11.1

cDNA Library ES3 - ATCC# 207025
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001362C:H11	945	RTA00000178AR.a.20.1
M00001362C:H11	945	100.E4.sp6:131473.Seq
M00001362C:H11	945	80.E4.sp6:130259.Seq
M00001362C:H11	945	180.C2.sp6:135940.Seq
M00001376B:G06	17732	RTA00000178AR.i.2.2
M00001376B:G06	17732	80.B5.sp6:130224.Seq
M00001387A:C05	2464	80.D6.sp6:130249.Seq
M00001387A:C05	2464	RTA00000178AF.n.18.1
M00001412B:B10	8551	RTA00000179AF.p.21.1
M00001412B:B10	8551	80.G7.sp6:130286.Seq
M00001415A:H06	13538	80.B8.sp6:130227.Seq
M00001415A:H06	13538	RTA00000180AF.a.24.1
M00001416B:H11	8847	80.C8.sp6:130239.Seq
M00001416B:H11	8847	RTA00000180AF.b.16.1
M00001429D:D07	40392	RTA00000180AF.j.8.1
M00001429D:D07	40392	80.H9.sp6:130300.Seq
M00001448D:H01	36313	80.A11.sp6:130218.Seq
M00001448D:H01	36313	RTA00000181AF.e.23.1

cDNA Library ES4 - ATCC# 207026
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001463C:B11	19	RTA00000182AF.b.7.1
M00001463C:B11	19	89.D1.sp6:130703.Seq
M00001470A:B10	1037	89.F2.sp6:130728.Seq
M00001470A:B10	1037	RTA00000121A.f.8.1
M00001497A:G02	2623	89.F3.sp6:130729.Seq
M00001497A:G02	2623	RTA00000183AF.a.6.1
M00001500A:E11	2623	RTA00000183AF.b.14.1
M00001500A:E11	2623	89.A4.sp6:130670.Seq
M00001501D:C02	9685	RTA00000183AF.c.11.1.Seq_THC109544
M00001501D:C02	9685	RTA00000183AF.c.11.1
M00001501D:C02	9685	89.C4.sp6:130694.Seq
M00001504C:H06	6974	89.F4.sp6:130730.Seq
M00001504C:H06	6974	RTA00000183AF.d.9.1
M00001504C:H06	6974	RTA00000183AF.d.9.1.Seq_THC223129
M00001504D:G06	6420	173.F5.SP6:134133.Seq
M00001504D:G06	6420	89.G4.sp6:130742.Seq
M00001504D:G06	6420	RTA00000183AF.d.11.1.Seq_THC226443
M00001504D:G06	6420	RTA00000183AF.d.11.1
M00001528A:C04	35555	89.B6.sp6:130684.Seq
M00001528A:C04	7337	RTA00000123A.b.17.1
M00001528A:C04	35555	184.A5.sp6:135530.Seq

cDNA Library ES5 - ATCC# 207027
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001537B:G07	3389	RTA00000183AF.m.19.1
M00001537B:G07	3389	89.A8.sp6:130674.Seq
M00001541A:D02	3765	89.C8.sp6:130698.Seq
M00001541A:D02	3765	RTA00000135A.d.1.1
M00001544B:B07	6974	89.A9.sp6:130675.Seq
M00001544B:B07	6974	RTA00000184AF.a.15.1
M00001546A:G11	1267	89.D9.sp6:130711.Seq
M00001546A:G11	1267	RTA00000125A.o.5.1
M00001549B:F06	4193	89.G9.sp6:130747.Seq
M00001549B:F06	4193	RTA00000184AF.e.13.1
M00001556A:F11	1577	173.C9.SP6:134101.Seq
M00001556A:F11	1577	89.F11.sp6:130737.Seq
M00001556A:F11	1577	RTA00000184AF.i.23.1
M00001556B:C08	4386	RTA00000184AF.j.4.1
M00001556B:C08	4386	89.H11.sp6:130761.Seq

cDNA Library ES6 - ATCC# 207028
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name

M00001563B:F06	102	RTA00000184AF.o.5.1
M00001563B:F06	102	90.B1.sp6:130871.Seq
M00001571C:H06	5749	90.E1.sp6:130907.Seq
M00001571C:H06	5749	RTA00000185AF.a.19.1
M00001594B:H04	260	90.D2.sp6:130896.Seq
M00001594B:H04	260	RTA00000185AR.i.12.2
M00001597C:H02	4837	90.E2.sp6:130908.Seq
M00001597C:H02	4837	RTA00000185AR.k.3.2
M00001624C:F01	4309	90.C4.sp6:130886.Seq
M00001624C:F01	4309	RTA00000186AF.e.22.1
M00001679A:A06	6660	90.F6.sp6:130924.Seq
M00001679A:A06	6660	122.B5.sp6:132089.Seq
M00001679A:A06	6660	RTA00000187AF.h.15.1
M00003759B:B09	697	90.G8.sp6:130938.Seq
M00003759B:B09	697	RTA00000188AF.d.6.1
M00003759B:B09	697	RTA00000188AF.d.6.1.Seq_THC178884
M00003844C:B11	6539	176.D9.sp6:134556.Seq
M00003844C:B11	6539	RTA00000189AF.d.22.1
M00003844C:B11	6539	90.B10.sp6:130880.Seq
M00003857A:G10	3389	90.A11.sp6:130869.Seq
M00003857A:G10	3389	RTA00000189AF.g.3.1

cDNA Library ES7 - ATCC# 207029
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00003914C:F05	3900	99.E1.sp6:131278.Seq
M00003914C:F05	3900	RTA00000190AF.g.13.1
M00003922A:E06	23255	RTA00000190AF.j.4.1
M00003922A:E06	23255	99.F1.sp6:131290.Seq
M00003922A:E06	23255	RTA00000190AF.j.4.1.Seq_THC228776
M00003983A:A05	9105	99.C3.sp6:131256.Seq
M00003983A:A05	9105	RTA00000191AF.a.21.2
M00004028D:A06	6124	RTA00000191AR.e.2.3
M00004028D:A06	6124	99.D3.sp6:131268.Seq
M00004031A:A12	9061	RTA00000191AR.e.11.2
M00004031A:A12	9061	RTA00000191AR.e.11.3
M00004087D:A01	6880	RTA00000191AF.m.20.1
M00004087D:A01	6880	99.A5.sp6:131234.Seq
M00004108A:E06	4937	99.E5.sp6:131282.Seq
M00004108A:E06	4937	RTA00000191AF.p.21.1
M00004114C:F11	13183	123.D5.sp6:132305.Seq
M00004114C:F11	13183	RTA00000192AF.a.24.1
M00004114C:F11	13183	99.G5.sp6:131306.Seq

cDNA Library ES8 - ATCC# 207030
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name

M00004146C:C11	5257	99.B6.sp6:131247.Seq
M00004146C:C11	5257	177.F5.sp6:134768.Seq
M00004146C:C11	5257	RTA00000192AF.f.3.1
M00004146C:C11	5257	RTA00000192AF.f.3.1.Seq_THC213833
M00004157C:A09	6455	RTA00000192AF.g.23.1
M00004157C:A09	6455	99.D6.sp6:131271.Seq
M00004157C:A09	6455	123.E7.sp6:132319.Seq
M00004172C:D08	11494	RTA00000192AF.j.6.1
M00004172C:D08	11494	99.G6.sp6:131307.Seq
M00004172C:D08	11494	177.E6.sp6:134757.Seq
M00004229B:F08	6455	RTA00000193AF.b.9.1
M00004229B:F08	6455	99.C8.sp6:131261.Seq

cDNA Library ES9 - ATCC# 207031
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001466A:E07	4275	RTA00000120A.j.14.1
M00001531A:H11		89.F6.sp6:130732.Seq
M00001531A:H11		RTA00000123A.g.19.1
M00001551A:B10	6268	79.G9.sp6:130096.Seq
M00001551A:B10	6268	184.C12.sp6:135561.Seq
M00001551A:B10	6268	RTA00000126A.o.23.1
M00001552A:B12	307	RTA00000136A.o.4.2
M00001552A:B12	307	79.C7.sp6:130046.Seq
M00001556A:H01	15855	RTA00000184AF.j.1.1
M00001586C:C05	4623	RTA00000185AF.f.4.1
M00001604A:B10	1399	79.G8.sp6:130095.Seq
M00001604A:B10	1399	RTA00000129A.o.10.1
M00003879B:C11	5345	RTA00000189AF.l.19.1
M00003879B:C11	5345	90.B12.sp6:130882.Seq

cDNA Library ES10 - ATCC#207032
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001358C:C06		RTA00000177AF.o.4.3
M00001388D:G05	5832	80.F6.sp6:130273.Seq
M00001388D:G05	5832	RTA00000178AF.o.23.1
M00001394A:F01	6583	RTA00000179AF.d.13.1
M00001394A:F01	6583	172.B8.sp6:133896.Seq
M00001394A:F01	6583	80.H6.sp6:130297.Seq
M00001429A:H04	2797	RTA00000180AF.i.19.1
M00001447A:G03	10717	RTA00000181AF.d.10.1
M00001448D:C09	8	80.H10.sp6:130301.Seq
M00001448D:C09	8	RTA00000181AF.e.17.1
M00001448D:C09	8	100.B11.sp6:131444.Seq
M00001454D:G03	689	RTA00000181AR.l.22.1

cDNA Library ES11 - ATCC#207033
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00003975A:G11	12439	RTA00000190AF.o.24.1
M00003978B:G05	5693	RTA00000190AF.p.17.2.Seq_THC173318
M00003978B:G05	5693	RTA00000190AF.p.17.2
M00004059A:D06	5417	RTA00000191AF.h.19.1
M00004068B:A01	3706	99.C4.sp6:131257.Seq
M00004068B:A01	3706	RTA00000191AF.i.17.2
M00004205D:F06		99.E7.sp6:131284.Seq
M00004205D:F06		177.G7.sp6:134782.Seq
M00004205D:F06		RTA00000192AF.o.11.1
M00004212B:C07	2379	RTA00000192AF.p.8.1
M00004223A:G10	16918	RTA00000193AF.a.16.1

cDNA Library ES12 - ATCC# 207034
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00004223B:D09	7899	RTA00000193AF.a.17.1
M00004249D:G12		RTA00000193AF.c.22.1
M00004251C:G07		RTA00000193AF.d.2.1
M00004372A:A03	2030	RTA00000193AF.m.20.1

cDNA Library ES13 - ATCC#207035
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001340B:A06	17062	80.A1.sp6:130208.Seq
M00001340B:A06	17062	RTA00000177AF.b.8.4
M00001340D:F10	11589	80.B1.sp6:130220.Seq
M00001340D:F10	11589	RTA00000177AF.b.17.4
M00001341A:E12	4443	80.C1.sp6:130232.Seq
M00001341A:E12	4443	RTA00000177AF.b.20.4
M00001342B:E06	39805	80.D1.sp6:130244.Seq
M00001342B:E06	39805	RTA00000177AF.c.21.3
M00001346A:F09	5007	RTA00000177AF.g.2.1
M00001346A:F09	5007	80.H1.sp6:130292.Seq
M00001346D:G06	5779	RTA00000177AF.g.14.3
M00001346D:G06	5779	RTA00000177AF.g.14.1
M00001348B:B04	16927	80.E2.sp6:130257.Seq
M00001348B:B04	16927	RTA00000177AF.h.9.3
M00001348B:G06	16985	RTA00000177AF.h.10.1
M00001348B:G06	16985	80.F2.sp6:130269.Seq
M00001349B:B08	3584	RTA00000177AF.h.20.1
M00001349B:B08	3584	80.G2.sp6:130281.Seq
M00001350A:H01	7187	100.C2.sp6:131447.Seq
M00001350A:H01	7187	80.A3.sp6:130210.Seq
M00001350A:H01	7187	RTA00000177AF.i.8.2

Clone Name	Cluster ID	Sequence Name
M00001352A:E02	16245	RTA00000177AF.k.9.3
M00001352A:E02	16245	172.D2.sp6:133914.Seq
M00001352A:E02	16245	80.D3.sp6:130246.Seq
M00001355B:G10	14391	RTA00000177AF.m.17.3
M00001355B:G10	14391	80.G3.sp6:130282.Seq
M00001355B:G10	14391	172.H3.sp6:133963.Seq
M00001355B:G10	14391	100.E3.sp6:131472.Seq
M00001361D:F08	2379	80.C4.sp6:130235.Seq
M00001361D:F08	2379	RTA00000178AF.a.6.1
M00001365C:C10	40132	RTA00000178AF.c.7.1
M00001365C:C10	40132	80.F4.sp6:130271.Seq
M00001368D:E03		80.G4.sp6:130283.Seq
M00001368D:E03		RTA00000178AF.d.20.1
M00001370A:C09	6867	80.H4.sp6:130295.Seq
M00001370A:C09	6867	RTA00000178AF.e.12.1
M00001371C:E09	7172	100.A5.sp6:131426.Seq
M00001371C:E09	7172	RTA00000178AF.f.9.1
M00001371C:E09	7172	80.A5.sp6:130212.Seq
M00001378B:B02	39833	80.C5.sp6:130236.Seq
M00001378B:B02	39833	RTA00000178AF.i.23.1
M00001379A:A05	1334	80.D5.sp6:130248.Seq
M00001379A:A05	1334	RTA00000178AF.j.7.1
M00001380D:B09	39886	RTA00000178AF.j.24.1
M00001380D:B09	39886	80.E5.sp6:130260.Seq
M00001381D:E06		80.F5.sp6:130272.Seq
M00001381D:E06		RTA00000178AF.k.16.1
M00001382C:A02	22979	80.G5.sp6:130284.Seq
M00001382C:A02	22979	RTA00000178AF.k.22.1
M00001384B:A11		80.B6.sp6:130225.Seq
M00001384B:A11		RTA00000178AF.m.13.1
M00001386C:B12	5178	80.C6.sp6:130237.Seq
M00001386C:B12	5178	RTA00000178AF.n.10.1
M00001387B:G03	7587	80.E6.sp6:130261.Seq
M00001387B:G03	7587	RTA00000178AF.n.24.1
M00001389A:C08	16269	RTA00000178AF.p.1.1
M00001389A:C08	16269	80.G6.sp6:130285.Seq
M00001396A:C03	4009	172.D8.sp6:133920.Seq
M00001396A:C03	4009	80.A7.sp6:130214.Seq
M00001396A:C03	4009	RTA00000179AF.e.20.1
M00001400B:H06		172.B9.sp6:133897.Seq
M00001400B:H06		80.B7.sp6:130226.Seq
M00001400B:H06		RTA00000179AF.j.13.1
M00001400B:H06		RTA00000179AF.j.13.1.Seq_THC105720
M00001402A:E08	39563	80.C7.sp6:130238.Seq
M00001402A:E08	39563	RTA00000179AF.k.20.1
M00001407B:D11	5556	RTA00000179AF.n.10.1

Clone Name	Cluster ID	Sequence Name
M00001407B:D11	5556	80.D7.sp6:130250.Seq
M00001410A:D07	7005	180.H5.sp6:136003.Seq
M00001410A:D07	7005	RTA00000179AF.o.22.1
M00001410A:D07	7005	80.F7.sp6:130274.Seq
M00001414A:B01		RTA00000180AF.a.9.1
M00001414A:B01		80.H7.sp6:130298.Seq
M00001414C:A07		80.A8.sp6:130215.Seq
M00001414C:A07		RTA00000180AF.a.11.1
M00001416A:H01	7674	79.C1.sp6:130040.Seq
M00001416A:H01	7674	RTA00000118A.g.9.1
M00001417A:E02	36393	RTA00000180AF.c.2.1
M00001417A:E02	36393	80.D8.sp6:130251.Seq
M00001423B:E07	15066	RTA00000180AF.e.24.1
M00001423B:E07	15066	80.H8.sp6:130299.Seq
M00001424B:G09	10470	80.A9.sp6:130216.Seq
M00001424B:G09	10470	RTA00000180AF.f.18.1
M00001425B:H08	22195	RTA00000180AF.g.7.1
M00001425B:H08	22195	80.B9.sp6:130228.Seq
M00001426B:D12		RTA00000180AF.g.22.1
M00001426B:D12		80.C9.sp6:130240.Seq
M00001426D:C08	4261	80.D9.sp6:130252.Seq
M00001426D:C08	4261	RTA00000180AF.h.5.1
M00001428A:H10	84182	100.G9.sp6:131502.Seq
M00001428A:H10	84182	RTA00000180AF.h.19.1
M00001428A:H10	84182	80.E9.sp6:130264.Seq
M00001449A:A12	5857	80.B11.sp6:130230.Seq
M00001449A:A12	5857	RTA00000118A.g.14.1
M00001449A:B12	41633	80.C11.sp6:130242.Seq
M00001449A:B12	41633	RTA00000118A.g.16.1
M00001449A:G10	36535	RTA00000181AF.f.5.1
M00001449A:G10	36535	80.D11.sp6:130254.Seq
M00001449A:G10	36535	100.D11.sp6:131468.Seq
M00001449C:D06	86110	RTA00000181AF.f.12.1
M00001449C:D06	86110	80.E11.sp6:130266.Seq
M00001450A:A02	39304	RTA00000118A.j.21.1.Seq_THC151859
M00001450A:A02	39304	RTA00000118A.j.21.1
M00001450A:A02	39304	79.F1.sp6:130076.Seq
M00001450A:A02	39304	180.G9.sp6:135995.Seq
M00001450A:A11	32663	80.F11.sp6:130278.Seq
M00001450A:A11	32663	RTA00000118A.l.8.1
M00001450A:B12	82498	100.F11.sp6:131492.Seq
M00001450A:B12	82498	RTA00000118A.m.10.1
M00001450A:B12	82498	79.G1.sp6:130088.Seq
M00001450A:D08	27250	80.G11.sp6:130290.Seq
M00001450A:D08	27250	180.B10.sp6:135936.Seq
M00001450A:D08	27250	RTA00000181AF.g.10.1

Clone Name	Cluster ID	Sequence Name
M00001452A:B04	84328	RTA00000118A.p.10.1
M00001452A:B04	84328	79.A2.sp6:130017.Seq
M00001452A:B12	86859	RTA00000118A.p.8.1
M00001452A:B12	86859	79.B2.sp6:130029.Seq
M00001452A:F05	85064	RTA00000131A.m.23.1
M00001452A:F05	85064	79.D2.sp6:130053.Seq
M00001452C:B06	16970	80.H11.sp6:130302.Seq
M00001452C:B06	16970	100.C12.sp6:131457.Seq
M00001452C:B06	16970	RTA00000181AR.i.18.2
M00001453A:E11	16130	80.A12.sp6:130219.Seq
M00001453A:E11	16130	100.D12.sp6:131469.Seq
M00001453A:E11	16130	RTA00000119A.c.13.1
M00001453C:F06	16653	80.B12.sp6:130231.Seq
M00001453C:F06	16653	RTA00000181AF.k.5.3
M00001454A:A09	83103	RTA00000119A.e.24.2
M00001454A:A09	83103	79.G2.sp6:130089.Seq
M00001454B:C12	7005	121.D1.sp6:131917.Seq
M00001454B:C12	7005	RTA00000181AF.k.24.1
M00001454B:C12	7005	80.C12.sp6:130243.Seq
M00001455B:E12	13072	80.F12.sp6:130279.Seq
M00001455B:E12	13072	RTA00000181AR.m.5.2
M00001460A:F06	2448	89.A1.sp6:130667.Seq
M00001460A:F06	2448	RTA00000119A.j.21.1
M00001461A:D06	1531	89.C1.sp6:130691.Seq
M00001461A:D06	1531	RTA00000119A.o.3.1
M00001465A:B11	10145	79.F3.sp6:130078.Seq
M00001465A:B11	10145	RTA00000120A.g.12.1
M00001467A:B07	38759	89.F1.sp6:130727.Seq
M00001467A:B07	38759	RTA00000120A.m.12.3
M00001467A:D04	39508	RTA00000120A.o.2.1
M00001467A:D04	39508	89.G1.sp6:130739.Seq
M00001467A:E10	39442	89.A2.sp6:130668.Seq
M00001467A:E10	39442	RTA00000120A.o.21.1
M00001468A:F05	7589	RTA00000120A.p.23.1
M00001468A:F05	7589	89.B2.sp6:130680.Seq
M00001469A:A01		RTA00000121A.c.10.1
M00001469A:A01		89.C2.sp6:130692.Seq
M00001469A:C10	12081	89.D2.sp6:130704.Seq
M00001469A:C10	12081	RTA00000133A.d.14.2
M00001469A:H12	19105	89.E2.sp6:130716.Seq
M00001469A:H12	19105	RTA00000133A.e.15.1
M00001470A:C04	39425	89.G2.sp6:130740.Seq
M00001470A:C04	39425	RTA00000133A.f.1.1
M00001471A:B01	39478	89.H2.sp6:130752.Seq
M00001471A:B01	39478	RTA00000133A.i.5.1
M00001487B:H06		RTA00000182AF.l.15.1

Clone Name	Cluster ID	Sequence Name
M00001487B:H06		89.B3.sp6:130681.Seq
M00001488B:F12		RTA00000182AF.l.20.1
M00001488B:F12		89.C3.sp6:130693.Seq
M00001494D:F06	7206	RTA00000182AF.o.15.1
M00001494D:F06	7206	89.E3.sp6:130717.Seq
M00001499B:A11	10539	RTA00000183AF.a.24.1
M00001499B:A11	10539	89.G3.sp6:130741.Seq
M00001499B:A11	10539	173.B5.SP6:134085.Seq
M00001500A:C05	5336	RTA00000183AF.b.13.1
M00001500A:C05	5336	89.H3.sp6:130753.Seq
M00001504A:E01		RTA00000183AF.c.24.1
M00001504A:E01		89.D4.sp6:130706.Seq
M00001504A:E01		RTA00000183AF.c.24.1.Seq_THC125912
M00001504C:A07	10185	RTA00000183AF.d.5.1
M00001504C:A07	10185	89.E4.sp6:130718.Seq
M00001505C:C05		89.H4.sp6:130754.Seq
M00001505C:C05		RTA00000183AF.e.1.1
M00001506D:A09		89.A5.sp6:130671.Seq
M00001506D:A09		RTA00000183AF.e.23.1
M00001506D:A09		121.G6.sp6:131958.Seq
M00001507A:H05	39168	RTA00000121A.l.10.1
M00001507A:H05	39168	89.B5.sp6:130683.Seq
M00001535A:F10	39423	79.C5.sp6:130044.Seq
M00001535A:F10	39423	RTA00000134A.k.22.1
M00001541A:H03	39174	79.E5.sp6:130068.Seq
M00001541A:H03	39174	RTA00000124A.n.13.1
M00001544A:G02	19829	79.H5.sp6:130104.Seq
M00001544A:G02	19829	RTA00000125A.h.24.4
M00001545A:D08	13864	RTA00000125A.m.9.1
M00001545A:D08	13864	79.B6.sp6:130033.Seq
M00001551A:F05	39180	RTA00000126A.n.8.2
M00001551A:F05	39180	79.A7.sp6:130022.Seq
M00001552A:D11	39458	RTA00000126A.p.15.2
M00001552A:D11	39458	79.D7.sp6:130058.Seq
M00001557A:F03	39490	RTA00000128A.b.4.1

cDNA Library ES14 - ATCC# 207036
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001511A:H06	39412	RTA00000133A.k.17.1
M00001511A:H06	39412	89.C5.sp6:130695.Seq
M00001512A:A09	39186	89.D5.sp6:130707.Seq
M00001512A:A09	39186	RTA00000121A.p.15.1
M00001512D:G09	3956	89.E5.sp6:130719.Seq
M00001512D:G09	3956	173.H5.SP6:134157.Seq
M00001512D:G09	3956	RTA00000183AF.g.3.1

Clone Name	Cluster ID	Sequence Name
M00001513B:G03		RTA00000183AF.g.9.1
M00001513B:G03		89.F5.sp6:130731.Seq
M00001513B:G03		RTA00000183AF.g.9.1.Seq_THC198280
M00001513C:E08	14364	RTA00000183AF.g.12.1
M00001513C:E08	14364	89.G5.sp6:130743.Seq
M00001514C:D11	40044	RTA00000183AF.g.22.1
M00001514C:D11	40044	RTA00000183AF.g.22.1.Seq_THC232899
M00001514C:D11	40044	89.H5.sp6:130755.Seq
M00001518C:B11	8952	89.A6.sp6:130672.Seq
M00001518C:B11	8952	RTA00000183AF.h.15.1
M00001528B:H04	8358	89.D6.sp6:130708.Seq
M00001528B:H04	8358	RTA00000183AF.i.5.1
M00001531A:D01	38085	RTA00000123A.e.15.1
M00001531A:D01	38085	89.E6.sp6:130720.Seq
M00001534A:C04	16921	RTA00000183AF.k.6.1
M00001534A:C04	16921	89.H6.sp6:130756.Seq
M00001534A:D09	5097	RTA00000134A.k.1.1
M00001534A:D09	5097	RTA00000134A.k.1.1.Seq_THC215869
M00001534C:A01	4119	RTA00000183AF.k.16.1
M00001534C:A01	4119	89.C7.sp6:130697.Seq
M00001535A:C06	20212	89.E7.sp6:130721.Seq
M00001535A:C06	20212	RTA00000134A.l.22.1.Seq_THC128232
M00001535A:C06	20212	RTA00000134A.l.22.1
M00001536A:B07	2696	RTA00000134A.m.13.1
M00001536A:B07	2696	89.F7.sp6:130733.Seq
M00001537A:F12	39420	89.H7.sp6:130757.Seq
M00001537A:F12	39420	RTA00000134A.o.23.1
M00001540A:D06	8286	89.B8.sp6:130686.Seq
M00001540A:D06	8286	RTA00000183AF.o.1.1
M00001542A:E06	39453	89.E8.sp6:130722.Seq
M00001542A:E06	39453	RTA00000135A.g.11.1
M00001544A:E06		RTA00000184AF.a.8.1
M00001544A:E06		173.G7.SP6:134147.Seq
M00001544A:E06		89.H8.sp6:130758.Seq
M00001545A:B02		89.B9.sp6:130687.Seq
M00001545A:B02		RTA00000135A.l.2.2
M00001548A:E10	5892	89.E9.sp6:130723.Seq
M00001548A:E10	5892	RTA00000184AF.d.11.1
M00001548A:E10	5892	RTA00000184AF.d.11.1.Seq_THC161896
M00001549C:E06	16347	89.H9.sp6:130759.Seq
M00001549C:E06	16347	RTA00000184AF.e.15.1
M00001550A:A03	7239	89.A10.sp6:130676.Seq
M00001550A:A03	7239	RTA00000126A.m.4.2
M00001550A:G01	5175	RTA00000184AF.f.3.1
M00001550A:G01	5175	89.B10.sp6:130688.Seq
M00001551A:G06	22390	RTA00000136A.j.13.1

Clone Name	Cluster ID	Sequence Name
M00001551A:G06	22390	89.C10.sp6:130700.Seq
M00001551C:G09	3266	RTA00000184AR.g.1.1
M00001551C:G09	3266	89.D10.sp6:130712.Seq
M00001553A:H06	8298	RTA00000127A.d.19.1
M00001553A:H06	8298	89.G10.sp6:130748.Seq
M00001553B:F12	4573	89.H10.sp6:130760.Seq
M00001553B:F12	4573	RTA00000184AF.h.9.1
M00001555A:B02	39539	RTA00000127A.i.21.1
M00001555A:B02	39539	89.B11.sp6:130689.Seq
M00001555A:C01	39195	89.C11.sp6:130701.Seq
M00001555A:C01	39195	RTA00000137A.c.16.1
M00001555D:G10	4561	RTA00000184AF.i.21.1
M00001555D:G10	4561	89.D11.sp6:130713.Seq
M00001556A:C09	9244	89.E11.sp6:130725.Seq
M00001556A:C09	9244	RTA00000127A.l.3.1
M00001556B:G02	11294	RTA00000184AF.j.6.1
M00001556B:G02	11294	89.A12.sp6:130678.Seq
M00001557B:H10	5192	173.E9.SP6:134125.Seq
M00001557B:H10	5192	RTA00000184AF.k.2.1
M00001557B:H10	5192	89.D12.sp6:130714.Seq
M00001557D:D09	8761	RTA00000184AF.k.12.1
M00001557D:D09	8761	89.E12.sp6:130726.Seq
M00001558B:H11	7514	RTA00000184AF.k.21.1
M00001558B:H11	7514	89.G12.sp6:130750.Seq
M00001559B:F01		89.H12.sp6:130762.Seq
M00001559B:F01		RTA00000184AF.l.11.1
M00001560D:F10	6558	90.A1.sp6:130859.Seq
M00001560D:F10	6558	RTA00000184AF.m.21.1
M00001566B:D11		RTA00000184AF.p.3.1
M00001566B:D11		90.D1.sp6:130895.Seq
M00001583D:A10	6293	RTA00000185AF.e.11.1
M00001583D:A10	6293	90.A2.sp6:130860.Seq
M00001590B:F03		RTA00000185AF.g.11.1
M00001590B:F03		90.C2.sp6:130884.Seq
M00001597D:C05	10470	RTA00000185AF.k.6.1
M00001597D:C05	10470	90.F2.sp6:130920.Seq
M00001598A:G03	16999	90.G2.sp6:130932.Seq
M00001598A:G03	16999	RTA00000185AF.k.9.1
M00001601A:D08	22794	RTA00000138A.b.5.1
M00001601A:D08	22794	90.H2.sp6:130944.Seq
M00001607A:E11	11465	RTA00000185AF.m.19.1
M00001607A:E11	11465	90.A3.sp6:130861.Seq
M00001608A:B03	7802	RTA00000185AF.n.5.1
M00001608A:B03	7802	90.B3.sp6:130873.Seq
M00001608B:E03	22155	RTA00000185AF.n.9.1
M00001608B:E03	22155	90.C3.sp6:130885.Seq

Clone Name	Cluster ID	Sequence Name
M00001608D:A11		RTA00000185AF.n.12.1
M00001608D:A11		90.D3.sp6:130897.Seq
M00001614C:F10	13157	RTA00000186AF.a.6.1
M00001614C:F10	13157	90.E3.sp6:130909.Seq
M00001617C:E02	17004	RTA00000186AF.b.21.1
M00001617C:E02	17004	90.F3.sp6:130921.Seq
M00001619C:F12	40314	90.G3.sp6:130933.Seq
M00001619C:F12	40314	RTA00000186AF.c.15.1
M00001621C:C08	40044	RTA00000186AF.d.1.1
M00001621C:C08	40044	RTA00000186AF.d.1.1.Seq_THC232899
M00001621C:C08	40044	90.H3.sp6:130945.Seq
M00001621C:C08	40044	122.E1.sp6:132121.Seq
M00001623D:F10	13913	RTA00000186AF.e.6.1
M00001623D:F10	13913	90.A4.sp6:130862.Seq
M00001632D:H07		RTA00000186AF.h.14.1.Seq_THC112525
M00001632D:H07		RTA00000186AF.h.14.1
M00001632D:H07		90.E4.sp6:130910.Seq
M00001632D:H07		176.A3.sp6:134514.Seq
M00001644C:B07	39171	RTA00000186AF.l.7.1
M00001644C:B07	39171	90.F4.sp6:130922.Seq
M00001644C:B07	39171	217.A12.sp6:139369.Seq
M00001645A:C12	19267	RTA00000186AF.l.12.1.Seq_THC178183
M00001645A:C12	19267	176.G3.sp6:134586.Seq
M00001645A:C12	19267	RTA00000186AF.l.12.1
M00001645A:C12	19267	90.G4.sp6:130934.Seq
M00001648C:A01	4665	90.H4.sp6:130946.Seq
M00001648C:A01	4665	RTA00000186AF.m.3.1
M00001657D:C03	23201	RTA00000187AF.a.14.1
M00001657D:C03	23201	90.B5.sp6:130875.Seq
M00001657D:F08	76760	90.C5.sp6:130887.Seq
M00001657D:F08	76760	RTA00000187AF.a.15.1
M00001662C:A09	23218	RTA00000187AR.c.5.2
M00001662C:A09	23218	90.D5.sp6:130899.Seq
M00001663A:E04	35702	90.E5.sp6:130911.Seq
M00001663A:E04	35702	RTA00000187AR.c.15.2
M00001669B:F02	6468	90.F5.sp6:130923.Seq
M00001669B:F02	6468	RTA00000187AF.d.15.1
M00001670C:H02	14367	90.G5.sp6:130935.Seq
M00001670C:H02	14367	RTA00000187AF.e.8.1
M00001673C:H02	7015	90.H5.sp6:130947.Seq
M00001673C:H02	7015	RTA00000187AF.f.18.1
M00001675A:C09	8773	RTA00000187AF.f.24.1
M00001675A:C09	8773	90.A6.sp6:130864.Seq
M00001675A:C09	8773	RTA00000187AF.f.24.1.Seq_THC220002
M00001676B:F05	11460	RTA00000187AF.g.12.1
M00001676B:F05	11460	90.B6.sp6:130876.Seq

Clone Name	Cluster ID	Sequence Name
M00001676B:F05	11460	219.F2.sp6:139035.Seq
M00001677D:A07	7570	90.D6.sp6:130900.Seq
M00001677D:A07	7570	RTA00000187AF.g.24.1
M00001677D:A07	7570	RTA00000187AF.g.24.1.Seq_THC168636
M00001678D:F12	4416	90.E6.sp6:130912.Seq
M00001678D:F12	4416	RTA00000187AF.h.13.1
M00001679A:F10	26875	RTA00000187AF.i.1.1
M00001679A:F10	26875	90.A7.sp6:130865.Seq
M00001679B:F01	6298	90.B7.sp6:130877.Seq
M00001679B:F01	6298	RTA00000187AR.i.10.2
M00001680D:F08	10539	90.F7.sp6:130925.Seq
M00001680D:F08	10539	219.F6.sp6:139039.Seq
M00001680D:F08	10539	RTA00000187AF.l.7.1
M00001682C:B12	17055	90.G7.sp6:130937.Seq
M00001682C:B12	17055	RTA00000187AF.m.3.1
M00001682C:B12	17055	176.D6.sp6:134553.Seq
M00001688C:F09	5382	90.A8.sp6:130866.Seq
M00001688C:F09	5382	RTA00000187AF.m.23.2
M00001693C:G01	4393	RTA00000187AF.n.17.1
M00001693C:G01	4393	90.B8.sp6:130878.Seq
M00001716D:H05	67252	RTA00000187AF.o.6.1
M00001716D:H05	67252	90.C8.sp6:130890.Seq
M00003741D:C09	40108	90.D8.sp6:130902.Seq
M00003741D:C09	40108	RTA00000187AF.o.24.1
M00003747D:C05	11476	RTA00000187AF.p.19.1
M00003747D:C05	11476	90.E8.sp6:130914.Seq
M00003747D:C05	11476	RTA00000187AF.p.19.1.Seq_THC108482
M00003747D:C05	11476	219.H8.sp6:139065.Seq
M00003754C:E09		90.F8.sp6:130926.Seq
M00003754C:E09		RTA00000188AF.b.12.1
M00003761D:A09		RTA00000188AF.d.11.1
M00003761D:A09		90.H8.sp6:130950.Seq
M00003761D:A09		RTA00000188AF.d.11.1.Seq_THC212094
M00003762C:B08	17076	RTA00000188AF.d.21.1.Seq_THC208760
M00003762C:B08	17076	90.A9.sp6:130867.Seq
M00003762C:B08	17076	RTA00000188AF.d.21.1
M00003763A:F06	3108	RTA00000188AF.d.24.1
M00003763A:F06	3108	90.B9.sp6:130879.Seq
M00003774C:A03	67907	RTA00000188AF.g.11.1.Seq_THC123222
M00003774C:A03	67907	RTA00000188AF.g.11.1
M00003774C:A03	67907	90.C9.sp6:130891.Seq
M00003784D:D12		RTA00000188AF.i.8.1
M00003784D:D12		90.D9.sp6:130903.Seq
M00003839A:D08	7798	RTA00000189AF.c.18.1
M00003839A:D08	7798	90.A10.sp6:130868.Seq
M00003851B:D08		90.D10.sp6:130904.Seq

Clone Name	Cluster ID	Sequence Name
M00003851B:D08		RTA00000189AF.f.7.1
M00003851B:D10	13595	90.E10.sp6:130916.Seq
M00003851B:D10	13595	RTA00000189AF.f.8.1
M00003853A:D04	5619	90.F10.sp6:130928.Seq
M00003853A:D04	5619	RTA00000189AF.f.17.1
M00003853A:F12	10515	90.G10.sp6:130940.Seq
M00003853A:F12	10515	RTA00000189AF.f.18.1
M00003856B:C02	4622	90.H10.sp6:130952.Seq
M00003856B:C02	4622	RTA00000189AF.g.1.1
M00003857A:H03	4718	90.B11.sp6:130881.Seq
M00003857A:H03	4718	RTA00000189AF.g.5.1.Seq_THC196102
M00003857A:H03	4718	RTA00000189AF.g.5.1

cDNA Library ES15 - ATCC# 207037
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00003867A:D10		90.C11.sp6:130893.Seq
M00003867A:D10		RTA00000189AF.h.17.1
M00003871C:E02	4573	RTA00000189AF.j.12.1
M00003875C:G07	8479	90.G11.sp6:130941.Seq
M00003875C:G07	8479	RTA00000189AF.j.22.1
M00003875D:D11		90.H11.sp6:130953.Seq
M00003875D:D11		RTA00000189AF.j.23.1
M00003876D:E12	7798	90.A12.sp6:130870.Seq
M00003876D:E12	7798	RTA00000189AF.k.12.1
M00003906C:E10	9285	90.H12.sp6:130954.Seq
M00003906C:E10	9285	RTA00000190AF.d.7.1
M00003907D:A09	39809	99.A1.sp6:131230.Seq
M00003907D:A09	39809	RTA00000190AF.e.3.1.Seq_THC150217
M00003907D:A09	39809	RTA00000190AF.e.3.1
M00003907D:H04	16317	99.B1.sp6:131242.Seq
M00003907D:H04	16317	RTA00000190AF.e.6.1
M00003909D:C03	8672	RTA00000190AF.f.11.1
M00003909D:C03	8672	99.C1.sp6:131254.Seq
M00003968B:F06	24488	RTA00000190AF.n.16.1
M00003968B:F06	24488	99.C2.sp6:131255.Seq
M00003970C:B09	40122	RTA00000190AF.n.23.1
M00003970C:B09	40122	RTA00000190AF.n.23.1.Seq_THC109227
M00003970C:B09	40122	99.D2.sp6:131267.Seq
M00003974D:E07	23210	RTA00000190AF.o.20.1
M00003974D:E07	23210	RTA00000190AF.o.20.1.Seq_THC207240
M00003974D:E07	23210	99.E2.sp6:131279.Seq
M00003974D:H02	23358	RTA00000190AF.o.21.1.Seq_THC207240
M00003974D:H02	23358	RTA00000190AF.o.21.1
M00003974D:H02	23358	99.F2.sp6:131291.Seq
M00003981A:E10	3430	99.A3.sp6:131232.Seq

Clone Name	Cluster ID	Sequence Name
M00003981A:E10	3430	RTA00000191AF.a.9.1
M00003982C:C02	2433	RTA00000191AF.a.15.2
M00003982C:C02	2433	99.B3.sp6:131244.Seq
M00003982C:C02	2433	RTA00000191AF.a.15.2.Seq_THC79498
M00004028D:C05	40073	RTA00000191AF.e.3.1
M00004028D:C05	40073	99.E3.sp6:131280.Seq
M00004035C:A07	37285	99.H3.sp6:131316.Seq
M00004035C:A07	37285	RTA00000191AF.f.11.1
M00004035D:B06	17036	RTA00000191AF.f.13.1
M00004035D:B06	17036	99.A4.sp6:131233.Seq
M00004072A:C03		RTA00000191AF.j.9.1
M00004072A:C03		99.D4.sp6:131269.Seq
M00004081C:D10	15069	99.F4.sp6:131293.Seq
M00004081C:D10	15069	RTA00000191AF.l.6.1
M00004086D:G06	9285	99.H4.sp6:131317.Seq
M00004086D:G06	9285	RTA00000191AF.m.18.1
M00004105C:A04	7221	99.D5.sp6:131270.Seq
M00004105C:A04	7221	RTA00000191AF.p.9.1
M00004171D:B03	4908	RTA00000192AF.j.2.1
M00004171D:B03	4908	99.F6.sp6:131295.Seq
M00004185C:C03	11443	RTA00000192AF.l.13.2
M00004185C:C03	11443	123.A8.sp6:132272.Seq
M00004185C:C03	11443	99.A7.sp6:131236.Seq
M00004191D:B11		RTA00000192AF.m.12.1
M00004191D:B11		99.B7.sp6:131248.Seq
M00004191D:B11		123.C8.sp6:132296.Seq
M00004197D:H01	8210	99.C7.sp6:131260.Seq
M00004197D:H01	8210	123.E8.sp6:132320.Seq
M00004197D:H01	8210	RTA00000192AF.n.13.1
M00004203B:C12	14311	99.D7.sp6:131272.Seq
M00004203B:C12	14311	RTA00000192AF.o.2.1
M00004214C:H05	11451	177.D8.sp6:134747.Seq
M00004214C:H05	11451	RTA00000192AF.p.17.1
M00004223D:E04	12971	RTA00000193AF.a.20.1
M00004223D:E04	12971	99.B8.sp6:131249.Seq
M00004269D:D06	4905	99.H8.sp6:131321.Seq
M00004269D:D06	4905	RTA00000193AF.e.14.1
M00004295D:F12	16921	99.D9.sp6:131274.Seq
M00004295D:F12	16921	RTA00000193AF.h.15.1
M00004296C:H07	13046	99.E9.sp6:131286.Seq
M00004296C:H07	13046	RTA00000193AF.h.19.1
M00004307C:A06	9457	RTA00000193AF.i.14.2
M00004307C:A06	9457	99.F9.sp6:131298.Seq
M00004307C:A06	9457	123.D11.sp6:132311.Seq
M00004312A:G03	26295	RTA00000193AF.i.24.2
M00004312A:G03	26295	99.G9.sp6:131310.Seq

Clone Name	Cluster ID	Sequence Name
M00004312A:G03	26295	RTA00000193AF.i.24.2.Seq_THC197345
M00004318C:D10	21847	RTA00000193AF.j.9.1
M00004318C:D10	21847	99.H9.sp6:131322.Seq
M00004359B:G02		RTA00000193AF.m.5.1.Seq_THC173318
M00004359B:G02		RTA00000193AF.m.5.1
M00004505D:F08		RTA00000194AF.b.19.1
M00004505D:F08		99.H10.sp6:131323.Seq
M00004692A:H08		99.B11.sp6:131252.Seq
M00004692A:H08		RTA00000194AF.c.24.1
M00004692A:H08		377.F4.sp6:141957.Seq
M00005180C:G03		RTA00000194AF.f.4.1

cDNA Library ES16 - ATCC#207038
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001346D:E03	6806	RTA00000177AF.g.13.3
M00001350A:B08		80.H2.sp6:130293.Seq
M00001350A:B08		RTA00000177AF.i.6.2
M00001357D:D11	4059	RTA00000177AF.n.18.3.Seq_THC123051
M00001357D:D11	4059	RTA00000177AF.n.18.3
M00001409C:D12	9577	RTA00000179AF.o.17.1
M00001409C:D12	9577	80.E7.sp6:130262.Seq
M00001418B:F03	9952	RTA00000180AF.c.20.1
M00001418B:F03	9952	RTA00000180AF.c.20.1.Seq_THC162284
M00001418B:F03	9952	80.E8.sp6:130263.Seq
M00001418D:B06	8526	RTA00000180AF.d.1.1
M00001421C:F01	9577	RTA00000180AF.d.23.1
M00001421C:F01	9577	80.G8.sp6:130287.Seq
M00001429B:A11	4635	RTA00000180AF.i.20.1
M00001432C:F06		RTA00000180AF.k.24.1
M00001439C:F08	40054	RTA00000180AF.p.10.1
M00001442C:D07	16731	RTA00000181AF.a.20.1
M00001442C:D07	16731	80.C10.sp6:130241.Seq
M00001443B:F01		80.D10.sp6:130253.Seq
M00001443B:F01		RTA00000181AF.b.7.1
M00001445A:F05	13532	80.E10.sp6:130265.Seq
M00001445A:F05	13532	RTA00000181AF.c.4.1
M00001446A:F05	7801	RTA00000181AF.c.21.1
M00001455A:E09	13238	RTA00000181AF.m.4.1
M00001455A:E09	13238	RTA00000181AF.m.4.1.Seq_THC140691
M00001460A:F12	39498	RTA00000119A.j.20.1
M00001481D:A05	7985	RTA00000182AR.j.2.1
M00001490B:C04	18699	RTA00000182AF.m.16.1
M00001490B:C04	18699	89.D3.sp6:130705.Seq
M00001500C:E04	9443	89.B4.sp6:130682.Seq
M00001500C:E04	9443	RTA00000183AF.c.1.1

cDNA Library ES16 - ATCC#207038
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001532B:A06	3990	89.G6.sp6:130744.Seq
M00001532B:A06	3990	RTA00000183AF.j.11.1
M00001534A:F09	5321	89.B7.sp6:130685.Seq
M00001534A:F09	5321	RTA00000183AF.k.8.1
M00001535A:B01	7665	RTA00000134A.l.19.1
M00001536A:C08	39392	89.G7.sp6:130745.Seq
M00001536A:C08	39392	RTA00000134A.m.16.1
M00001541A:F07	22085	RTA00000135A.e.5.2
M00001542B:B01		RTA00000183AF.p.4.1
M00001542B:B01		89.F8.sp6:130734.Seq
M00001544A:E03	12170	RTA00000125A.h.18.4
M00001545A:C03	19255	RTA00000135A.m.18.1
M00001545A:C03	19255	184.B10.sp6:135547.Seq
M00001545A:C03	19255	89.C9.sp6:130699.Seq
M00001548A:H09	1058	RTA00000126A.e.20.3.Seq_THC217534
M00001548A:H09	1058	RTA00000126A.e.20.3
M00001548A:H09	1058	79.F6.sp6:130081.Seq
M00001549A:B02	4015	RTA00000136A.e.12.1
M00001549A:B02	4015	79.G6.sp6:130093.Seq
M00001549A:D08	10944	RTA00000126A.h.17.2
M00001552B:D04	5708	RTA00000184AF.g.12.1
M00001552B:D04	5708	89.E10.sp6:130724.Seq
M00001552D:A01		89.F10.sp6:130736.Seq
M00001552D:A01		RTA00000184AF.g.22.1
M00001553D:D10	22814	RTA00000184AF.h.14.1
M00001553D:D10	22814	89.A11.sp6:130677.Seq
M00001558A:H05		RTA00000128A.c.20.1
M00001558A:H05		89.F12.sp6:130738.Seq
M00001561A:C05	39486	RTA00000128A.m.22.2
M00001561A:C05	39486	79.B8.sp6:130035.Seq
M00001564A:B12	5053	RTA00000184AF.o.12.1
M00001578B:E04	23001	RTA00000185AF.c.24.1
M00001579D:C03	6539	90.G1.sp6:130931.Seq
M00001579D:C03	6539	173.A12.SP6:134080.Seq
M00001579D:C03	6539	RTA00000185AF.d.11.1
M00001582D:F05		RTA00000185AF.d.24.1
M00001587A:B11	39380	RTA00000129A.e.24.1
M00001587A:B11	39380	79.E8.sp6:130071.Seq
M00001604A:F05	39391	RTA00000138A.c.3.1
M00001604A:F05	39391	79.A9.sp6:130024.Seq
M00001624A:B06	3277	RTA00000138A.l.5.1
M00001624A:B06	3277	217.E1.sp6:139406.Seq
M00001624A:B06	3277	90.B4.sp6:130874.Seq
M00001630B:H09	5214	90.D4.sp6:130898.Seq

cDNA Library ES16 - ATCC#207038
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00001630B:H09	5214	122.C2.sp6:132098.Seq
M00001630B:H09	5214	RTA00000186AF.g.11.1
M00001651A:H01		RTA00000186AF.n.7.1
M00001651A:H01		90.A5.sp6:130863.Seq
M00001677C:E10	14627	RTA00000187AF.g.23.1
M00001679C:F01	78091	90.C7.sp6:130889.Seq
M00001679C:F01	78091	RTA00000187AF.j.6.1
M00001679C:F01	78091	176.G5.sp6:134588.Seq
M00001686A:E06	4622	RTA00000187AF.m.15.2
M00003796C:D05	5619	RTA00000188AF.l.9.1.Seq_THC167845
M00003796C:D05	5619	RTA00000188AF.l.9.1
M00003826B:A06	11350	RTA00000189AF.a.24.2
M00003826B:A06	11350	90.F9.sp6:130927.Seq
M00003833A:E05	21877	RTA00000189AF.b.21.1
M00003837D:A01	7899	90.H9.sp6:130951.Seq
M00003837D:A01	7899	RTA00000189AF.c.10.1
M00003846B:D06	6874	RTA00000189AF.e.9.1
M00003846B:D06	6874	90.C10.sp6:130892.Seq
M00003879B:D10	31587	RTA00000189AF.l.20.1
M00003879B:D10	31587	90.C12.sp6:130894.Seq
M00003879D:A02	14507	90.D12.sp6:130906.Seq
M00003879D:A02	14507	RTA00000189AR.l.23.2
M00003891C:H09		90.G12.sp6:130942.Seq
M00003891C:H09		RTA00000189AF.p.8.1
M00003912B:D01	12532	99.D1.sp6:131266.Seq
M00003912B:D01	12532	RTA00000190AF.g.2.1
M00004072B:B05	17036	RTA00000191AF.j.10.1
M00004081C:D12	14391	RTA00000191AF.l.7.1
M00004111D:A08	6874	RTA00000192AF.a.14.1
M00004111D:A08	6874	99.F5.sp6:131294.Seq
M00004121B:G01		177.H4.sp6:134791.Seq
M00004121B:G01		99.H5.sp6:131318.Seq
M00004121B:G01		RTA00000192AF.c.2.1
M00004138B:H02	13272	99.A6.sp6:131235.Seq
M00004138B:H02	13272	RTA00000192AF.e.3.1
M00004151D:B08	16977	RTA00000192AF.g.3.1
M00004169C:C12	5319	99.E6.sp6:131283.Seq
M00004169C:C12	5319	RTA00000192AF.i.12.1
M00004169C:C12	5319	123.F7.sp6:132331.Seq
M00004183C:D07	16392	RTA00000192AF.l.1.1
M00004183C:D07	16392	RTA00000192AF.l.1.1.Seq_THC202071
M00004230B:C07	7212	RTA00000193AF.b.14.1
M00004230B:C07	7212	99.D8.sp6:131273.Seq
M00004249D:F10		RTA00000193AF.c.21.1.Seq_THC222602

cDNA Library ES16 - ATCC#207038
Deposit Date - December 22, 1998

Clone Name	Cluster ID	Sequence Name
M00004249D:F10		RTA00000193AF.c.21.1
M00004275C:C11	16914	99.A9.sp6:131238.Seq
M00004275C:C11	16914	RTA00000193AF.f.5.1
M00004283B:A04	14286	RTA00000193AF.f.22.1
M00004285B:E08	56020	RTA00000193AF.g.2.1
M00004327B:H04		RTA00000193AF.j.20.1
M00004377C:F05	2102	RTA00000193AF.n.7.1
M00004384C:D02		RTA00000193AF.n.15.1
M00004384C:D02		RTA00000193AF.n.15.1.Seq_THC215687
M00004461A:B08		RTA00000194AR.a.10.2
M00004461A:B09		RTA00000194AF.a.11.1
M00004691D:A05		RTA00000194AF.c.23.1
M00004896A:C07		RTA00000194AF.d.13.1

The above material has been deposited with the American Type Culture Collection, Rockville, Maryland, under the accession number indicated. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for purposes of Patent Procedure. The deposit will be maintained for a period of 30 years following issuance of this patent, or for the enforceable life of the patent, whichever is greater. Upon issuance of the patent, the deposit will be available to the public from the ATCC without restriction.

This deposit is provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby.

Retrieval of Individual Clones from Deposit of Pooled Clones

Where the ATCC deposit is composed of a pool of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones were then deposited as a pool of equal mixtures in the composite deposit. Particular clones can be

obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a T_m of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C). Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

Example 14: Source of Biological Materials and Overview of Novel Polynucleotides Expressed by the Biological Materials

Human colon cancer cell line Km12L4-A (Morika, W. A. K. et al., *Cancer Research* (1988) *48*:6863) was used to construct a cDNA library from mRNA isolated from the cells. As described in the above overview, a total of 4,693 sequences expressed by the Km12L4-A cell line were isolated and analyzed; most sequences were about 275-300 nucleotides in length. The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B_2 surgical specimen (Morikawa *et al. Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman *et al. Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling *et al. Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as a model cell line for the study of colon cancer (see, *e.g.*, Moriakawa *et al.*, *supra*; Radinsky *et al. Clin. Cancer Res.* (1995) *1*:19; Yeatman *et al.*, (1995) *supra*; Yeatman *et al. Clin. Exp. Metastasis* (1996) *14*:246).

The sequences were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol.* 266:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis Techniques" Adams *et al.*, eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie *et al. Comput. Chem.* (1993) 17:191). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats. Masking resulted in the elimination of 43 sequences. The remaining sequences were then used in a BLASTN vs. Genbank search with search parameters of greater than 70% overlap, 99% identity, and a p value of less than 1×10^{-40}, which search resulted in the discarding of 1,432 sequences. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database search: (1) unknown (no hits in the Genbank search), (2) weak similarity (greater than

45% identity and p value of less than 1 x 10^{-5}), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1 x 10^{-5}). This search resulted in discard of 98 sequences as having greater than 70% overlap, greater than 99% identity, and p value of less than 1 x 10^{-40}.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search resulted in discard of 1771 sequences (sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1 x 10^{-40}; sequences with a p value of less than 1 x 10^{-65} when compared to a database sequence of human origin were also excluded). Second, a BLASTN vs. Patent GeneSeq database resulted in discard of 15 sequences (greater than 99% identity; p value less than 1 x 10^{-40}; greater than 99% overlap).

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1 x 10^{-111} in relation to a database sequence of human origin were specifically excluded. The final result provided the 2502 sequences listed in the accompanying Sequence Listing. The Sequence Listing is arranged beginning with sequences with no similarity to any sequence in a database searched, and ending with sequences with the greatest similarity. Each identified polynucleotide represents sequence from at least a partial mRNA transcript. Polynucleotides that were determined to be novel were assigned a sequence identification number.

The novel polynucleotides were assigned sequence identification numbers SEQ ID NOS:845-3346. The DNA sequences corresponding to the novel polynucleotides are provided in the Sequence Listing. The majority of the sequences are presented in the Sequence Listing in the 5' to 3' direction. A small number of sequences are listed in the Sequence Listing in the 5' to 3' direction but the sequence as written is actually 3' to 5'. These sequences are readily identified with the designation "AR" in the Sequence Name in Table 17 (inserted before the claims). The sequences correctly listed in the 5' to 3' direction in the Sequence Listing are designated "AF." Table 17 provides: 1) the SEQ ID NO assigned to each sequence for use in the present specification; 2) the filing date of the U.S. priority application in which the sequence was first filed; 3) the SEQ ID NO assigned to the sequence in the priority application; 4) the sequence name used as an internal

identifier of the sequence; 5) the name assigned to the clone from which the sequence was isolated; and 6) the number of the cluster to which the sequence is assigned (Cluster ID; where the cluster ID is 0, the sequence was not assigned to any cluster

Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOS: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene. In addition, some sequences are identified with multiple SEQ ID NOS, since these sequences were present in more than one filing. For example, SEQ ID NO:931 and SEQ ID NO:1844 represent the same sequence.

In order to confirm the sequences of SEQ ID NOS:845-3346, inserts of the clones corresponding to these polynucleotides were re-sequenced. These "validation" sequences are provided in SEQ ID NOS:3347-5106. Of these validation sequences, SEQ ID NOS:3384, 4389, 4407, 5355, 5570, and 5593 are not true validation sequences. Instead, SEQ ID NOS: 4389, 5355, 5570, and 5593 represent "placeholder" sequences, *i.e.*, sequences that were inserted into the Sequence Listing only to prevent renumbering of the subsequent sequences during generation of the Sequence Listing. Thus, reference to "SEQ ID NOS:845-6096," "SEQ ID NOS:845-5950," or other ranges of SEQ ID NOS that include these placeholder sequences should be read to exclude SEQ ID NOS: 4389, 5355, 5570, and 5593.

The validation sequences were often longer than the original polynucleotide sequences they validate, and thus often provide additional sequence information. Validation sequences can be correlated with the original sequences they validate by referring to Table 17. For example, validation sequences of many SEQ ID NOS share the clone name of the sequence that they validate.

Example 15: Results of Public Database Search to Identify Function of Gene Products

SEQ ID NOS:845-3346, as well as the validation sequences were translated in all three reading frames to determine the best alignment with the individual sequences. These amino acid sequences and nucleotide sequences are referred, generally, as query sequences, which are aligned with the individual sequences. Query and individual sequences were aligned using the BLAST programs, available over the world wide web site of the NCBI.

Again the sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above in Example 1.

Table 18 (inserted before the claims) shows the results of the alignments. Table 18 refers to each sequence by its SEQ ID NO:, the accession numbers and descriptions of nearest neighbors from the Genbank and Non-Redundant Protein searches, and the p values of the search results.

For each of "SEQ ID NOS:845-5950," the best alignment to a protein or DNA sequence is included in Table 18. The activity of the polypeptide encoded by "SEQ ID NOS: 845-5950" is the same or similar to the nearest neighbor reported in Table 18. The accession number of the nearest neighbor is reported, providing a reference to the activities exhibited by the nearest neighbor. The search program and database used for the alignment also are indicated as well as a calculation of the p value.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence of "SEQ ID NOS: 845-5950." The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of "SEQ ID NOS: 845-5950."

"SEQ ID NOS: 845-5950" and the translations thereof may be human homologs of known genes of other species or novel allelic variants of known human genes. In such cases, these new human sequences are suitable as diagnostics or therapeutics. As diagnostics, the human sequences "SEQ ID NOS: 845-5950" exhibit greater specificity in detecting and differentiating human cell lines and types than homologs of other species. The human polypeptides encoded by "SEQ ID NOS:845-5950" are likely to be less immunogenic when administered to humans than homologs from other species. Further, on administration to humans, the polypeptides encoded by "SEQ ID NOS: 845-5950" can show greater specificity or can be better regulated by other human proteins than are homologs from other species.

Example 16: Members of Protein Families

The validation sequences ("SEQ ID NOS:3347-5950") were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide

belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 19, inserted prior to claims). Thus the invention encompasses fragments, fusions, and variants of such polynucleotides that retain biological activity associated with the protein family and/or functional domain identified herein.

Start and stop indicate the position within the individual sequences that align with the query sequence having the indicated SEQ ID NO. The direction (Dir) indicates the orientation of the query sequence with respect to the individual sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Some polynucleotides exhibited multiple profile hits because, for example, the particular sequence contains overlapping profile regions, and/or the sequence contains two different functional domains. These profile hits are described in more detail below. The acronyms used in Table 19 are provided in parentheses following the full name of the protein family or functional domain to which they refer.

Table 19 Polynucleotides encoding gene products of a protein family or having a known functional domain(s).

SEQ ID NO:	Validation Sequence	Biological Activity (Profile)	Start	Stop	Score	Direction
4764	393.E10.sp6:148957	7tm_1	531	710	9520	for
3511	172.F10.sp6:133946	7tm_2	45	724	8708	rev
3602	177.C6.sp6:134733	7tm_2	41	697	9828	rev
3777	184.C7.sp6:135556	7tm_2	3	834	8987	for
3973	121.E12.sp6:131940	7tm_2	245	1324	9550	rev
4209	172.A7.sp6:133883	7tm_2	94	761	8743	rev
4262	123.F9.sp6:132333	7tm_2	203	585	8785	rev
4263	123.F9.sp6:132333	7tm_2	203	585	8785	rev
4441	394.G2.sp6:149165	7tm_2	73	793	9209	for
4492	370.C5.sp6:141726	7tm_2	76	770	9269	for
4530	370.B1.sp6:141710	7tm_2	89	662	8791	for
4539	368.A12.sp6:141322	7tm_2	121	719	9015	rev
4540	368.A12.sp6:141322	7tm_2	121	719	9015	rev
5016	219.C10.sp6:139007	7tm_2	46	774	11394	rev
5060	368.D11.sp6:141357	7tm_2	66	775	9384	rev
5072	368.A11.sp6:141321	7tm_2	7	1079	9097	for
5285	99.F7.sp6:131296	7tm_2	534	1265	10956	rev
5286	99.F7.sp6:131296	7tm_2	534	1265	10956	rev

SEQ ID NO:	Validation Sequence	Biological Activity (Profile)	Start	Stop	Score	Direction
5326	100.D2.sp6:131459	7tm_2	122	1404	9296	rev
5339	395.B12.sp6:149307	7tm_2	79	1432	10427	rev
5369	90.B4.sp6:130874	7tm_2	4	691	9435	for
5460	100.D5.sp6:131462	7tm_2	655	1349	9255	for
5497	100.D7.sp6:131464	7tm_2	357	1346	11461	rev
5498	100.D7.sp6:131464	7tm_2	357	1346	11461	rev
5502	100.H6.sp6:131511	7tm_2	119	1035	10001	rev
5503	100.G6.sp6:131499	7tm_2	363	1188	9901	rev
5504	100.F6.sp6:131487	7tm_2	50	1127	8799	for
5505	100.F6.sp6:131487	7tm_2	50	1127	8799	for
5554	367.H9.sp6:141210	7tm_2	143	1266	11883	rev
5599	370.F4.sp6:141761	7tm_2	78	704	8942	for
5700	367.H11.sp6:141212	7tm_2	176	1227	9975	rev
5729	123.E10.sp6:132322	7tm_2	210	691	9071	rev
5744	123.E10.sp6:132322	7tm_2	210	691	9071	rev
5745	123.E10.sp6:132322	7tm_2	210	691	9071	rev
3500	176.H11.sp6:134606	ANK	207	290	4450	for
3399	180.C9.sp6:135947	asp	156	670	6710	for
4476	368.H11.sp6:141405	asp	136	1226	6880	rev
5049	368.B5.sp6:141327	asp	309	806	6073	for
5095	369.D6.sp6:141546	asp	434	1332	6263	rev
5097	396.F9.sp6:149544	asp	97	1106	5999	rev
5105	216.G10.sp6:139247	asp	74	703	6188	rev
5209	122.H12.sp6:132168	asp	152	1040	6183	rev
5342	80.H6.sp6:130297	asp	61	418	5944	rev
5508	172.E5.sp6:133929	asp	219	976	6434	for
5562	185.D9.sp6:135762	asp	31	872	5944	rev
5577	185.D9.sp6:135762	asp	31	872	5944	rev
5590	176.B10.sp6:134533	asp	253	1446	6079	rev
5666	177.F3.sp6:134766	asp	0	894	6336	rev
5698	184.F11.sp6:135596	asp	61	737	6416	rev
5700	367.H11.sp6:141212	asp	81	1187	6182	rev
5773	180.E6.sp6:135968	asp	81	706	6150	for
5775	180.E6.sp6:135968	asp	81	706	6150	for
3567	180.F2.sp6:135976	ATPases	135	627	11664	for
3686	217.H11.sp6:139452	ATPases	2	701	5972	for
3863	216.B1.sp6:139178	ATPases	170	616	6150	for
3890	121.B8.sp6:131900	ATPases	13	635	5867	rev
4034	80.D2.sp6:130245	ATPases	13	386	6068	for
4134	176.C6.sp6:134541	ATPases	85	579	5883	for
4514	369.C10.sp6:141538	ATPases	329	730	6206	for
4842	394.H8.sp6:149183	ATPases	21	571	5954	rev
4963	218.F11.sp6:138852	ATPases	313	816	6057	for
5003	219.A7.sp6:138980	ATPases	88	662	6145	for
5067	368.F9.sp6:141379	ATPases	178	648	5937	for
5228	181.G11.sp6:135354	ATPases	362	769	5900	rev
5317	369.B4.sp6:141520	ATPases	4	412	14130	for
5384	218.C8.sp6:138813	ATPases	12	576	5782	rev

SEQ ID NO:	Validation Sequence	Biological Activity (Profile)	Start	Stop	Score	Direction
5404	404.G6.sp6:162933	ATPases	86	605	6001	rev
5533	367.H8.sp6:141209	ATPases	17	476	5905	rev
5629	184.E5.sp6:135578	ATPases	184	632	5943	for
5636	184.C6.sp6:135555	ATPases	333	813	5773	for
5691	184.B11.sp6:135548	ATPases	14	498	6140	for
5885	377.C1.sp6:141918	ATPases	4	655	5933	for
4248	176.F10.sp6:134581	Bcl-2	69	356	16419	for
4880	367.F5.sp6:141182	bromodomain	40	210	8810	for
5333	369.D3.sp6:141543	bromodomain	63	230	10270	for
4252	172.E1.sp6:133925	BZIP	146	298	4066	for
4795	393.G5.sp6:148976	BZIP	116	304	5931	for
5694	172.E9.sp6:133933	BZIP	91	260	4366	for
4462	370.B12.sp6:141721	cyclin	118	324	8980	for
4739	395.G6.sp6:149361	cyclin	11	281	6930	for
5380	395.G8.sp6:149363	cyclin	12	279	5950	for
5299	99.F5.sp6:131294	Cys-protease	72	348	18479	for
5528	180.D1.sp6:135951	Cys-protease	38	992	10103	rev
5532	180.D1.sp6:135951	Cys-protease	38	992	10103	rev
5645	177.E4.sp6:134755	Cys-protease	48	326	19999	for
5503	100.G6.sp6:131499	DAG_PE_bind	605	702	6290	rev
5665	377.C8.sp6:141925	Dead_box_helic	172	828	7867	rev
5927	216.A1.sp6:139166	Dead_box_helic	44	589	26532	for
3578	177.G4.sp6:134779	EFhand	79	153	3780	for
3737	185.A1.sp6:135718	EFhand	287	358	2580	rev
4619	377.A5.sp6:141898	EFhand	477	563	3010	for
4900	367.B7.sp6:141136	EFhand	225	272	2500	rev
4996	218.B10.sp6:138803	EFhand	40	114	2640	rev
4997	218.B10.sp6:138803	EFhand	40	114	2640	rev
4998	218.C10.sp6:138815	EFhand	39	113	2640	rev
5749	393.H12.sp6:148995	EFhand	145	231	4640	for
5787	219.A9.sp6:138982	EFhand	685	750	2550	rev
3693	218.B5.sp6:138798	Ets_Nterm	340	531	10400	for
3572	180.A2.sp6:135916	FNtypeII	291	423	6400	rev
3862	216.C1.sp6:139190	FNtypeII	501	634	6460	for
5340	218.G1.sp6:138854	FNtypeII	20	141	6180	rev
5758	393.H8.sp6:148991	FNtypeII	448	576	6110	for
3348	181.C3.sp6:135298	G-alpha	66	715	8084	rev
4134	176.C6.sp6:134541	G-alpha	62	690	9062	for
5132	121.B4.sp6:131896	G-alpha	46	447	21415	for
5288	217.D12.sp6:139405	G-alpha	15	702	40404	for
5406	404.B7.sp6:162874	G-alpha	120	682	8424	for
3347	180.A11.sp6:135925	helicase_C	165	479	4494	for
5313	369.C4.sp6:141532	helicase_C	559	756	3732	rev
5864	185.D12.sp6:135765	helicase_C	381	534	5000	for
5085	396.H8.sp6:149567	homeobox	80	230	5170	for
3394	180.E5.sp6:135967	mkk	342	612	5791	for
4251	172.F1.sp6:133937	mkk	94	669	5688	rev
4295	123.A2.sp6:132266	mkk	26	378	7889	for

SEQ ID NO:	Validation Sequence	Biological Activity (Profile)	Start	Stop	Score	Direction
4444	394.B3.sp6:149106	mkk	32	782	9544	for
4490	370.H4.sp6:141785	mkk	18	307	9394	for
4524	369.G11.sp6:141587	mkk	182	725	5375	for
5019	219.H10.sp6:139067	mkk	280	723	15454	for
5049	368.B5.sp6:141327	mkk	249	725	5502	for
5122	181.C9.sp6:135304	mkk	168	880	5551	rev
5166	121.F6.sp6:131946	mkk	111	730	5399	for
5621	177.E2.sp6:134753	mkk	288	636	5720	rev
5326	100.D2.sp6:131459	PDEase	849	1195	5945	for
3422	181.H11.sp6:135366	protkinase	116	710	5531	for
3556	177.G7.sp6:134782	protkinase	6	511	5445	for
3679	218.C1.sp6:138806	protkinase	127	747	5492	for
3687	218.E1.sp6:138830	protkinase	64	726	5592	rev
3815	217.F4.sp6:139421	protkinase	83	702	5818	rev
3853	217.A4.sp6:139361	protkinase	57	682	5395	rev
3928	121.E2.sp6:131930	protkinase	69	658	5593	rev
4070	100.D8.sp6:131465	protkinase	174	620	5453	for
4118	100.C3.sp6:131448	protkinase	228	736	5616	for
4200	172.B5.sp6:133893	protkinase	148	715	5381	for
4221	172.B6.sp6:133894	protkinase	119	775	5616	for
4295	123.A2.sp6:132266	protkinase	24	384	9797	for
4444	394.B3.sp6:149106	protkinase	357	780	11395	for
4479	377.G11.sp6:141976	protkinase	117	739	5992	for
4490	370.H4.sp6:141785	protkinase	24	275	8338	for
4509	370.F2.sp6:141759	protkinase	33	800	5658	for
4513	369.B10.sp6:141526	protkinase	1	482	5504	rev
4544	369.D2.sp6:141542	protkinase	28	661	5428	for
4554	369.G6.sp6:141582	protkinase	71	631	5751	for
4635	396.C11.sp6:149510	protkinase	27	709	5793	rev
4749	393.H7.sp6:148990	protkinase	88	680	5470	rev
4763	393.D10.sp6:148945	protkinase	72	594	5617	for
4888	367.G4.sp6:141193	protkinase	30	699	5439	for
4916	368.B2.sp6:141324	protkinase	44	800	5556	for
4961	218.D11.sp6:138828	protkinase	38	781	6423	for
5019	219.H10.sp6:139067	protkinase	277	717	15720	for
5217	216.E5.sp6:139218	protkinase	115	710	5537	for
5413	100.C10.sp6:131455	protkinase	56	783	5556	rev
5599	370.F4.sp6:141761	protkinase	39	803	5635	for
5604	370.F3.sp6:141760	protkinase	188	775	5771	for
5651	184.H3.sp6:135612	protkinase	23	699	5515	for
5903	180.B5.sp6:135931	protkinase	182	671	5718	rev
5946	393.F4.sp6:148963	protkinase	28	650	5345	for
4515	369.D10.sp6:141550	ras	12	332	9802	for
4780	393.A3.sp6:148902	Thioredox	0	263	5887	rev
4771	393.F11.sp6:148970	TNFR_c6	151	261	6445	for
3800	184.E10.sp6:135583	transmembrane4	19	483	8339	rev
3825	217.E6.sp6:139411	transmembrane4	83	728	8417	for
4680	396.C9.sp6:149508	transmembrane4	300	924	9444	rev

SEQ ID NO:	Validation Sequence	Biological Activity (Profile)	Start	Stop	Score	Direction
4882	367.A6.sp6:141123	transmembrane4	32	495	8407	rev
5208	123.A1.sp6:132265	transmembrane4	1289	1548	8114	rev
5250	122.C1.sp6:132097	transmembrane4	6	535	8122	for
5275	122.E4.sp6:132124	transmembrane4	10	530	8829	for
5285	99.F7.sp6:131296	transmembrane4	613	1253	9443	rev
5286	99.F7.sp6:131296	transmembrane4	613	1253	9443	rev
5497	100.D7.sp6:131464	transmembrane4	335	1207	8255	rev
5498	100.D7.sp6:131464	transmembrane4	335	1207	8255	rev
5554	367.H9.sp6:141210	transmembrane4	398	1130	8352	rev
5788	180.H7.sp6:136005	transmembrane4	356	983	8356	rev
4225	176.D9.sp6:134556	trypsin	164	764	9670	rev
5528	180.D1.sp6:135951	trypsin	371	1229	10479	rev
5532	180.D1.sp6:135951	trypsin	371	1229	10479	rev
3598	177.H6.sp6:134793	WD_domain	345	437	6510	for
3890	121.B8.sp6:131900	WD_domain	98	193	6400	for
4071	100.B10.sp6:131443	WD_domain	544	642	6590	for
5087	121.A8.sp6:131888	WD_domain	93	188	6400	for
5890	185.F10.sp6:135787	WD_domain	382	480	5880	for
3973	121.E12.sp6:131940	Wnt_dev_sign	101	821	12160	rev
4017	99.G6.sp6:131307	Wnt_dev_sign	49	880	12334	rev
4234	176.C9.sp6:134544	Wnt_dev_sign	249	854	11038	rev
4235	176.C9.sp6:134544	Wnt_dev_sign	249	854	11038	rev
4500	370.G6.sp6:141775	Wnt_dev_sign	211	785	11490	rev
4680	396.C9.sp6:149508	Wnt_dev_sign	282	1017	12318	rev
5097	396.F9.sp6:149544	Wnt_dev_sign	482	1298	11217	rev
5174	122.A2.sp6:132074	Wnt_dev_sign	94	933	12383	rev
5203	123.B2.sp6:132278	Wnt_dev_sign	538	1435	11785	for
5208	123.A1.sp6:132265	Wnt_dev_sign	760	1544	12660	rev
5219	122.G10.sp6:132154	Wnt_dev_sign	29	884	11603	rev
5229	122.A2.sp6:132074	Wnt_dev_sign	94	933	12383	rev
5253	121.F12.sp6:131952	Wnt_dev_sign	9	734	11167	rev
5285	99.F7.sp6:131296	Wnt_dev_sign	560	1399	13749	rev
5286	99.F7.sp6:131296	Wnt_dev_sign	560	1399	13749	rev
5379	395.F10.sp6:149353	Wnt_dev_sign	100	907	11535	rev
5430	123.A4.sp6:132268	Wnt_dev_sign	80	1122	11249	rev
5449	404.D5.sp6:162896	Wnt_dev_sign	31	816	11304	rev
5497	100.D7.sp6:131464	Wnt_dev_sign	467	1314	11882	rev
5498	100.D7.sp6:131464	Wnt_dev_sign	467	1314	11882	rev
5509	177.B11.sp6:134726	Wnt_dev_sign	137	1266	12708	rev
5512	177.B11.sp6:134726	Wnt_dev_sign	137	1266	12708	rev
5526	177.B11.sp6:134726	Wnt_dev_sign	137	1266	12708	rev
5554	367.H9.sp6:141210	Wnt_dev_sign	692	1481	12886	rev
5562	185.D9.sp6:135762	Wnt_dev_sign	129	890	11145	rev
5568	377.D2.sp6:141931	Wnt_dev_sign	400	1227	11044	rev
5577	185.D9.sp6:135762	Wnt_dev_sign	129	890	11145	rev
5700	367.H11.sp6:141212	Wnt_dev_sign	295	1669	13366	rev
5710	377.D4.sp6:141933	Wnt_dev_sign	549	1380	14522	rev
5769	219.B12.sp6:138997	Wnt_dev_sign	312	1214	13188	rev

SEQ ID NO:	Validation Sequence	Biological Activity (Profile)	Start	Stop	Score	Direction
5803	219.B12.sp6:138997	Wnt_dev_sign	312	1214	13188	rev
4253	172.D1.sp6:133913	Y_phosphatase	476	804	6932	for
4262	123.F9.sp6:132333	Y_phosphatase	28	439	6096	rev
4263	123.F9.sp6:132333	Y_phosphatase	28	439	6096	rev
4501	370.H6.sp6:141787	Y_phosphatase	148	554	6481	for
4648	404.B10.sp6:162877	Y_phosphatase	104	466	6446	rev
4650	404.D10.sp6:162901	Y_phosphatase	9	614	6516	for
4818	395.F2.sp6:149345	Y_phosphatase	164	645	6093	rev
5082	121.E9.sp6:131937	Y_phosphatase	240	777	6147	rev
5107	216.F10.sp6:139235	Y_phosphatase	21	504	6342	for
5187	122.E9.sp6:132129	Y_phosphatase	381	807	6036	rev
5207	123.B1.sp6:132277	Y_phosphatase	61	510	6229	rev
5278	219.F4.sp6:139037	Y_phosphatase	2	261	10353	for
5317	369.B4.sp6:141520	Y_phosphatase	231	768	6110	rev
5473	404.E11.sp6:162914	Y_phosphatase	580	920	6005	rev
5938	217.A3.sp6:139360	Y_phosphatase	263	622	6222	rev
3582	177.A6.sp6:134709	Zincfing_C2H2	65	127	4380	for
3604	177.A6.sp6:134709	Zincfing_C2H2	65	127	4380	for
3676	218.B2.sp6:138795	Zincfing_C2H2	94	156	4940	for
4580	377.H8.sp6:141985	Zincfing_C2H2	495	557	4850	for
4606	377.G2.sp6:141967	Zincfing_C2H2	52	114	4380	for
4607	377.G2.sp6:141967	Zincfing_C2H2	52	114	4380	for
5638	377.G4.sp6:141969	Zincfing_C2H2	247	308	3930	for
5934	185.C4.sp6:135745	Zincfing_C2H2	238	300	4540	for
4618	377.E4.sp6:141945	Zincfing_C3HC4	128	244	9335	for
5321	181.E3.sp6:135322	Zincfing_C3HC4	321	445	8221	for

a) Seven Transmembrane Integral Membrane Proteins -- Rhodopsin Family (7tm_1). Several of the validation sequences, and thus their corresponding sequence within SEQ ID NOS:845-3346, correspond to a sequence encoding a polypeptide that is a member of the seven transmembrane receptor rhodopsin family. G-protein coupled receptors of the seven transmembrane rhodopsin family (also called R7G) are an extensive group of hormones, neurotransmitters, and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins (Strosberg A.D. *Eur. J. Biochem.* (1991) *196*:1, Kerlavage A.R. *Curr. Opin. Struct. Biol.* (1991) *1*:394, Probst, et al., *DNA Cell Biol.* (1992) *11*:1, Savarese, et al., Biochem. J. (1992) *283*:1. The receptors that are currently known to belong to this family are: 1) 5-hydroxytryptamine (serotonin) 1A to 1F, 2A to 2C, 4, 5A, 5B, 6 and 7 (Branchek T., *Curr. Biol.* (1993) *3*:315); 2) acetylcholine, muscarinic-type, M1 to M5; 3) adenosine A1, A2A, A2B and A3 (Stiles G.L. *J. Biol. Chem.* (1992) *267*:6451; 4) adrenergic alpha-1A to -1C; alpha-2A to -2D; beta-1 to -3

(Friell T. et al., *Trends Neurosci.* (1988) *11*:321); 5) angiotensin II types I and II; 6) bombesin subtypes 3 and 4; 7) bradykinin B1 and B2; 8) c3a and C5a anaphylatoxin; 9) cannabinoid CB1 and CB2; 10) chemokines C-C CC-CKR-1 to CC-CKR-8; 11) Chemokines C-X-C CXC-CKR-1 to CXC-CKR-4; 12) Cholecystokinin-A and cholecystokinin-B/gastrin Dopamine D1 to D5 (Stevens C.F., *Curr. Biol.* (1991) *1*:20); 13) Endothelin ET-a and ET-b (Sakurai T. et al., *Trends Pharmacol. Sci.* (1992) *13*:103-107); 14) fMet-Leu-Phe (fMLP) (Nformyl peptide); 15) Follicle stimulating hormone (FSH-R); 16) Galanin; 17) Gastrin-releasing peptide (GRP-R); 18) Gonadotropin-releasing hormone (GNRH-R); 19) Histamine H1 and H2 (gastric receptor I); 20) Lutropin-choriogonadotropic hormone (LSH-R) (Salesse R., et al., *Biochimie* (1991) *73*:109); 21) Melanocortin MC1R to MC5R; 22) Melatonin; 23) Neuromedin B (NMB-R); 24) Neuromedin K (NK-3R); 25) Neuropeptide Y types 1 to 6; 26) Neurotensin (NT-R); 27) Octopamine (tyramine), from insects; 28) Odorants (Lancet D., et al., *Curr. Biol.* (1993)*3*:668; 29) Opioids delta-, kappa- and mu-types (Uhl G.R., et al., *Trends Neurosci.* (1994) *17*:89; 30) Oxytocin (OT-R); 31) Platelet activating factor (PAF-R); 32) Prostacyclin; 33) Prostaglandin D2; 34) Prostaglandin E2, EP1 to EP4 subtypes; 35) Prostaglandin F2; 36) Purinoreceptors (ATP) (Barnard E.A., et al., *Trends Pharmacol. Sci.* (1994)*15*:67; 37); Somatostatin types 1 to 5; 38) Substance-K (NK-2R); Substance-P (NK-1R); 39) Thrombin; 40) Thromboxane A2; 41) Thyrotropin (TSH-R) (Salesse R., et al., *Biochimie* (1991) *73*:109); 42) Thyrotropin releasing factor (TRH-R); 42) Vasopressin V1a, V1b and V2; 43) Visual pigments (opsins and rhodopsin) (Applebury M.L., et al., *Vision Res.* (1986) *26*:1881; 44) Proto-oncogene mas; 45) A number of orphan receptors (whose ligand is not known) from mammals and birds; 46) Caenorhabditis elegans putative receptors C06G4.5, C38C10.1, C43C3.2; 47) T27D1.3 and ZC84.4; 48) Three putative receptors encoded in the genome of cytomegalovirus: US27, US28, and UL33; and 49) ECRF3, a putative receptor encoded in the genome of herpesvirus saimiri.

The structure of these receptors is thought to be identical. They have seven hydrophobic regions, each of which most probably spans the membrane. The N-terminus is located on the extracellular side of the membrane and is often glycosylated, while the C-terminus is cytoplasmic and generally phosphorylated. Three extracellular loops alternate with three intracellular loops to link the seven transmembrane regions. Most, but not all of these receptors, lack a signal peptide. The most conserved parts of these proteins are the

transmembrane regions and the first two cytoplasmic loops. A conserved acidic-Arg-aromatic triplet is present in the N-terminal extremity of the second cytoplasmic loop (Attwood T.K., Eliopoulos E.E., Findlay J.B.C. *Gene* (1991) *98*:153-159) and could be implicated in the interaction with G proteins.

b) Seven Transmembrane Integral Membrane Proteins -- Secretin Family (7tm_2). Several of the validation sequences, and thus their corresponding sequence in the sequence listing, correspond to a sequence encoding a polypeptide that is a member of the seven transmembrane receptor secretin family. A number of peptide hormones bind to G-protein coupled receptors that, while structurally similar to the majority of G-protein coupled receptors (R7G) (see profile for 7 transmembrane receptors (rhodopsin family), do not show any similarity at the level of their sequence, thus new family whose current known members (Jueppner et al. *Science* (1991) *254*:1024; Hamann et al. *Genomics* (1996) *32*:144).are: 1) calcitonin receptor, 2) calcitonin gene-related peptide receptor; 3) corticotropin releasing factor receptor types 1 and 2; 4) gastric inhibitory polypeptide receptor; 5) glucagon receptor; 6) glucagon-like peptide 1 receptor; 7) growth hormone-releasing hormone receptor; 7) parathyroid hormone / parathyroid hormone-related peptide types 1 and 2; 8) pituitary adenylate cyclase activating polypeptide receptor; 9) secretin receptor; 10) vasoactive intestinal peptide receptor types 1 and 2; 10) insects diuretic hormone receptor; 11) Caenorhabditis elegans putative receptor C13B9.4; 12) Caenorhabditis elegans putative receptor ZK643.3; 13) human leucocyte CD97 (which contains 3 EGF-like domains in its N-terminal section); 14) human cell surface glycoprotein EMR1 (which contains 6 EGF-like domains in it N-terminal section); and 15) mouse cell surface glycoprotein F4/80 (which contains 7 EGF-like domains in its N-terminal section). All of 1) through 10) are coupled to G-proteins which activate both adenylyl cyclase and the phosphatidylinositol-calcium pathway.

Like classical R7G the secretin family of 7 transmembrane proteins contain seven transmembrane regions. Their N-terminus is located on the extracellular side of the membrane and potentially glycosylated, while their C-terminus is cytoplasmic. But apart from these topological similarities they do not share any region of sequence similarity and are therefore probably not evolutionary related.

Every receptor in the 7 transmember secretin family is encoded on multiple exons, and several of these functionally distinct products. The N-terminal extracellular domain of

these receptors contains five conserved cysteines residues that may be involved in disulfide bonds, with a consensus pattern in the region that spans the first three cysteines. One of the most highly conserved regions spans the C-terminal part of the last transmembrane region and the beginning of the adjacent intracellular region. This second region is used as a second signature pattern.

c) Ank Repeats (ANK). The ankyrin motif is a 33 amino acid sequence named after the protein ankyrin which has 24 tandem 33-amino-acid motifs. Ank repeats were originally identified in the cell-cycle-control protein cdc10 (Breeden *et al., Nature* (1987) *329*:651). Proteins containing ankyrin repeats include ankyrin, myotropin, I-kappaB proteins, cell cycle protein cdc10, the Notch receptor (Matsuno *et al., Development* (1997) *124(21)*:4265); G9a (or BAT8) of the class III region of the major histocompatibility complex (Biochem J. 290:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, and SW16. The functions of the ankyrin repeats are compatible with a role in protein-protein interactions (Bork, *Proteins* (1993) *17(4)*:363; Lambert and Bennet, *Eur. J. Biochem.* (1993) *211*:1; Kerr *et al., Current Op. Cell Biol.* (1992) *4*:496; Bennet *et al., J. Biol. Chem.* (1980) *255*:6424).

The 90 kD N-terminal domain of ankyrin contains a series of 24 33-amino-acid ank repeats. (Lux *et al., Nature* (1990) *344*:36-42, Lambert *et al., PNAS USA* (1990) *87*:1730.) The 24 ank repeats form four folded subdomains of 6 repeats each. These four repeat subdomains mediate interactions with at least 7 different families of membrane proteins. Ankyrin contains two separate binding sites for anion exchanger dimers. One site utilizes repeat subdomain two (repeats 7-12) and the other requires both repeat subdomains 3 and 4 (repeats 13-24). Since the anion exchangers exist in dimers, ankyrin binds 4 anion exchangers at the same time (Michaely and Bennett, *J. Biol. Chem.* (1995) *270(37)*:22050). The repeat motifs are involved in ankyrin interaction with tubulin, spectrin, and other membrane proteins. (Lux *et al., Nature* (1990) *344*:36.)

The Rel/NF-kappaB/Dorsal family of transcription factors have activity that is controlled by sequestration in the cytoplasm in association with inhibitory proteins referred to as I-kappaB. (Gilmore, *Cell* (1990) *62*:841; Nolan and Baltimore, *Curr Opin Genet Dev.* (1992) *2*:211; Baeuerle, *Biochim Biophys Acta* (1991) *1072*:63; Schmitz *et al., Trends Cell Biol.* (1991) *1*:130.) I-kappaB proteins contain 5 to 8 copies of 33 amino acid ankyrin repeats and certain NF-kappaB/rel proteins are also regulated by cis-acting ankyrin repeat

containing domains including p105NF-kappaB which contains a series of ankyrin repeats (Diehl and Hannink, *J. Virol.* (1993) *67(12)*:7161). The I-kappaBs and Cactus (also containing ankyrin repeats) inhibit activators through differential interactions with the Rel-homology domain. The gene family includes proto-oncogenes, thus broadly implicating I-kappaB in the control of both normal gene expression and the aberrant gene expression that makes cells cancerous. (Nolan and Baltimore, *Curr Opin Genet Dev.* (1992) *2(2)*:211-220). In the case of rel/NF-kappaB and pp40/I-kappaB(, both the ankyrin repeats and the carboxy-terminal domain are required for inhibiting DNA-binding activity and direct association of pp40/I-kappaB(with rel/NF-kappaB protein. The ankyrin repeats and the carboxy-terminal of pp40/I-kappaB(form a structure that associates with the rel homology domain to inhibit DNA binding activity (Inoue *et al.*, *PNAS USA* (1992) *89*:4333).

The 4 ankyrin repeats in the amino terminus of the transcription factor subunit GABP□ are required for its interaction with the GABP□ subunit to form a functional high affinity DNA-binding protein. These repeats can be crosslinked to DNA when GABP is bound to its target sequence. (Thompson *et al.*, *Science* (1991) *253*:762; LaMarco *et al.*, *Science* (1991) *253*:789). Myotrophin, a 12.5 kDa protein having a key role in the initiation of cardiac hypertrophy, comprises ankyrin repeats. The ankyrin repeats are characteristic of a hairpin-like protruding tip followed by a helix-turn-helix motif. The V-shaped helix-turn-helix of the repeats stack sequentially in bundles and are stabilized by compact hydrophobic cores, whereas the protruding tips are less ordered.

d) Eukaryotic Aspartyl Proteases (asp). Several of the validation sequences correspond to a sequence encoding a novel eukaryotic aspartyl protease. Aspartyl proteases, known as acid proteases, (EC 3.4.23.-) are a widely distributed family of proteolytic enzymes (Foltmann B., *Essays Biochem.* (1981) *17*:52; Davies D.R., *Annu. Rev. Biophys. Chem.* (1990) *19*:189; Rao J.K.M., *et al.*, *Biochemistry* (1991) *30*:4663) known to exist in vertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue. The two domains most probably evolved from the duplication of an ancestral gene encoding a primordial domain. Currently known eukaryotic aspartyl proteases include: 1) Vertebrate gastric pepsins A and C (also known as gastricsin); 2) Vertebrate chymosin (rennin), involved in digestion and used for making cheese; 3) Vertebrate lysosomal cathepsins D (EC 3.4.23.5) and E (EC

3.4.23.34); 4) Mammalian renin (EC 3.4.23.15) whose function is to generate angiotensin I from angiotensinogen in the plasma; 5) Fungal proteases such as aspergillopepsin A (EC 3.4.23.18), candidapepsin (EC 3.4.23.24), mucoropepsin (EC 3.4.23.23) (mucor rennin), endothiapepsin (EC 3.4.23.22), polyporopepsin (EC 3.4.23.29), and rhizopuspepsin (EC 3.4.23.21); and 6) Yeast saccharopepsin (EC 3.4.23.25) (proteinase A) (gene PEP4). PEP4 is implicated in posttranslational regulation of vacuolar hydrolases; 7) Yeast barrierpepsin (EC 3.4.23.35) (gene BAR1); a protease that cleaves alpha-factor and thus acts as an antagonist of the mating pheromone; and 8) Fission yeast sxa1 which is involved in degrading or processing the mating pheromones.

Most retroviruses and some plant viruses, such as badnaviruses, encode for an aspartyl protease which is an homodimer of a chain of about 95 to 125 amino acids. In most retroviruses, the protease is encoded as a segment of a polyprotein which is cleaved during the maturation process of the virus. It is generally part of the pol polyprotein and, more rarely, of the gag polyprotein. Because the sequence around the two aspartates of eukaryotic aspartyl proteases and around the single active site of the viral proteases is conserved, a single signature pattern can be used to identify members of both groups of proteases.

e) ATPases Associated with Various Cellular Activities (ATPases). Several of the validation sequences, correspond to a sequence that encodes a novel member of the "ATPases Associated with diverse cellular Activities" (AAA) protein family. The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids that contains an ATP-binding site (Froehlich *et al., J. Cell Biol.* (1991) *114*:443; Erdmann *et al. Cell* (1991) *64*:499; Peters *et al., EMBO J.* (1990) *9*:1757; Kunau *et al., Biochimie* (1993) *75*:209-224; Confalonieri *et al., BioEssays* (1995) *17*:639; http://yeamob.pci.chemie.uni-tuebingen.de/AAA/Description.html). The proteins that belong to this family either contain one or two AAA domains.

Proteins containing two AAA domains include: 1) Mammalian and drosophila NSF (N-ethylmaleimide-sensitive fusion protein) and the fungal homolog, SEC18, which are involved in intracellular transport between the endoplasmic reticulum and Golgi, as well as between different Golgi cisternae; 2) Mammalian transitional endoplasmic reticulum ATPase (previously known as p97 or VCP), which is involved in the transfer of membranes from the endoplasmic reticulum to the golgi apparatus. This ATPase forms a

ring-shaped homooligomer composed of six subunits. The yeast homolog, CDC48, plays a role in spindle pole proliferation; 3) Yeast protein PAS1 essential for peroxisome assembly and the related protein PAS1 from Pichia pastoris; 4) Yeast protein AFG2; 5) Sulfolobus acidocaldarius protein SAV and Halobacterium salinarium cdcH, which may be part of a transduction pathway connecting light to cell division.

Proteins containing a single AAA domain include: 1) Escherichia coli and other bacteria ftsH (or hflB) protein. FtsH is an ATP-dependent zinc metallopeptidase that degrades the heat-shock sigma-32 factor, and is an integral membrane protein with a large cytoplasmic C-terminal domain that contain both the AAA and the protease domains; 2) Yeast protein YME1, a protein important for maintaining the integrity of the mitochondrial compartment. YME1 is also a zinc-dependent protease; 3) Yeast protein AFG3 (or YTA10). This protein also contains an AAA domain followed by a zinc-dependent protease domain; 4) Subunits from regulatory complex of the 26S proteasome (Hilt *et al., Trends Biochem. Sci.* (1996) *21*:96), which is involved in the ATP-dependent degradation of ubiquitinated proteins, which subunits include: a) Mammalian 4 and homologs in other higher eukaryotes, in yeast (gene YTA5) and fission yeast (gene mts2); b) Mammalian 6 (TBP7) and homologs in other higher eukaryotes and in yeast (gene YTA2); c) Mammalian subunit 7 (MSS1) and homologs in other higher eukaryotes and in yeast (gene CIM5 or YTA3); d) Mammalian subunit 8 (P45) and homologs in other higher eukaryotes and in yeast (SUG1 or CIM3 or TBY1) and fission yeast (gene let1); e) Other probable subunits include human TBP1, which influences HIV gene expression by interacting with the virus tat transactivator protein, and yeast YTA1 and YTA6; 5) Yeast protein BCS1, a mitochondrial protein essential for the expression of the Rieske iron-sulfur protein; 6) Yeast protein MSP1, a protein involved in intramitochondrial sorting of proteins; 7) Yeast protein PAS8, and the corresponding proteins PAS5 from Pichia pastoris and PAY4 from Yarrowia lipolytica; 8) Mouse protein SKD1 and its fission yeast homolog (SpAC2G11.06); 9) Caenorhabditis elegans meiotic spindle formation protein mei-1; 10) Yeast protein SAP1' 11) Yeast protein YTA7; and 12) Mycobacterium leprae hypothetical protein A2126A.

In general, the AAA domains in these proteins act as ATP-dependent protein clamps(Confalonieri *et al.* (1995) *BioEssays 17*:639). In addition to the ATP-binding 'A' and 'B' motifs, which are located in the N-terminal half of this domain, thereis a highly

conserved region located in the central part of the domain which was used in the development of the signature pattern.

f) Bcl-2 family (Bcl-2). SEQ ID NO:4248, and thus the corresponding sequence it validates, represents a polynucleotide encoding an apoptosis regulator protein of the Bcl-2 family. Active cell suicide (apoptosis) is induced by events such as growth factor withdrawal and toxins. It is controlled by regulators, which have either an inhibitory effect on programmed cell death (anti-apoptotic) or block the protective effect of inhibitors (pro-apoptotic) (Vaux, 1993, Curr. Biol. 3:877-878, and White, 1996, Genes Dev. 10:2859-2869). Many viruses have found a way of countering defensive apoptosis by encoding their own anti-apoptosis genes, preventing their target cells from dying prematurely.

All proteins belonging to the Bcl-2 family (Reed et al., 1996, Adv. Exp. Med. Biol. 406:99-112) contain either a BH1, BH2, BH3, or BH4 domain. All anti-apoptotic proteins contain BH1 and BH2 domains; some of them contain an additional N-terminal BH4 domain (Bcl-2, Bcl-x(L), Bcl-w), which is never seen in pro-apoptotic proteins, except for Bcl-x(S). On the other hand, all pro-apoptotic proteins contain a BH3 domain (except for Bad) necessary for dimerization with other proteins of Bcl-2 family and crucial for their killing activity; some of them also contain BH1 and BH2 domains (Bax, Bak). The BH3 domain is also present in some anti-apoptotic protein, such as Bcl-2 or Bcl-x(L). Proteins that are known to contain these domains are listed below.

1. Vertebrate protein Bcl-2. Bcl-2 blocks apoptosis; it prolongs the survival of hematopoietic cells in the absence of required growth factors and also in the presence of various stimuli inducing cellular death. Two isoforms of bcl-2 (alpha and beta) are generated by alternative splicing. Bcl-2 is expressed in a wide range of tissues at various times during development. It forms heterodimers with the Bax proteins.

2. Vertebrate protein Bcl-x. Two isoforms of Bcl-x (Bcl-x(L) and Bcl-x(S)) are generated by alternative splicing. While the longer product (Bcl-x(L)) can protect a growth-factor-dependent cell line from apoptosis, the shorter form blocks the protective effect of Bcl-2 and Bcl-x(L) and acts as an anti-anti-apoptosis protein.

3. Mammalian protein Bax. Bax blocks the anti-apoptosis ability of Bcl-2 with which it forms heterodimers. There is no evidence that Bax has any activity in the absence of Bcl-2. Three isoforms of bax (alpha, beta and gamma) are generated by alternative splicing.

4. Mammalian protein Bak, which promotes cell death and counteracts the protection from apoptosis provided by Bcl-2.

5. Mammalian protein Bcl-w, which promotes cell survival.

6. Mammalian protein bad, which promotes cell death, and counteracts the protection from apoptosis provided by Bcl-x(L), but not that of Bcl-2.

7. Human protein Bik, which promotes cell death, but cannot counteract the protection from apoptosis provided by Bcl-2.

8. Mouse protein Bid, which induces caspases and apoptosis, and counteracts the protection from apoptosis provided by Bcl-2.

9. Human induced myeloid leukemia cell differentiation protein MCL1. MCL1 is probably involved in programming of differentiation and concomitant maintenance of viability but not proliferation. Its expression increases early during phorbol ester induced differentiation in myeloid leukemia cell line ML-1.

10. Mouse hemopoietic-specific early response protein A1.

11. Mammalian activator of apoptosis Harakiri (Inohara et al., 1997, EMBO J. 16:1686-1694) (also known as neuronal death protein Dp5). This is a small protein of 92 residues that activates apoptosis. It contains a BH3 domain, but no BH1, BH2 or BH4 domains.

The following consensus patterns have been developed for the four BH domains:

g) Bromodomain (bromodomain). Some SEQ ID NOS represent polynucleotides encoding a polypeptide having a bromodomain region (Haynes et al., 1992, Nucleic Acids Res. 20:2693-2603, Tamkun et al., 1992, Cell 68:561-572, and Tamkun, 1995, Curr. Opin. Genet. Dev. 5:473-477), which is a conserved region of about 70 amino acids found in the following proteins: 1) Higher eukaryotes transcription initiation factor TFIID 250 Kd subunit (TBP-associated factor p250) (gene CCG1); P250 is associated with the TFIID TATA-box binding protein and seems essential for progression of the G1 phase of the cell cycle. 2) Human RING3, a protein of unknown function encoded in the MHC class II locus; 3) Mammalian CREB-binding protein (CBP), which mediates cAMP-gene regulation by binding specifically to phosphorylated CREB protein; 4) Mammalian homologs of brahma, including three brahma-like human: SNF2a(hBRM), SNF2b, and BRG1; 5) Human BS69, a protein that binds to adenovirus E1A and inhibits E1A

transactivation; 6) Human peregrin (or Br140).

The bromodomain is thought to be involved in protein-protein interactions and may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation.

h) Basic Region Plus Leucine Zipper Transcription Factors (BZIP). Some SEQ ID NOS, and thus the corresponding sequences these sequences validate, represent polynucleotides encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization. Members of the family include transcription factor AP-1, which binds selectively to enhancer elements in the cis control regions of SV40 and metallothionein IIA. AP-1, also known as c-jun, is the cellular homolog of the avian sarcoma virus 17 (ASV17) oncogene v-jun.

Other members of this protein family include jun-B and jun-D, probable transcription factors that are highly similar to jun/AP-1; the fos protein, a proto-oncogene that forms a non-covalent dimer with c-jun; the fos-related proteins fra-1, and fos B; and mammalian cAMP response element (CRE) binding proteins CREB, CREM, ATF-1, ATF-3, ATF-4, ATF-5, ATF-6 and LRF-1.

i) Cyclins (cyclin). Some SEQ ID NOS represent polynucleotides encoding cyclins, and SEQ ID NO:899 and 900, respectively, show the corresponding full-length polynucleotides. SEQ ID NO:901 and 902 show, respectively, the translations of SEQ ID NO:899 and 900. Cyclins (Nurse, 1990, Nature 344:503-508; Norbury et al., 1991, Curr. Biol. 1:23-24; and Lew et al., 1992, Trends Cell Biol. 2:77-81) are eukaryotic proteins that play an active role in controlling nuclear cell division cycles. There are two main groups of cyclins. G2/M cyclins are essential for the control of the cell cycle at the G2/M (mitosis) transition. G2/M cyclins accumulate steadily during G2 and are abruptly destroyed as cells exit from mitosis (at the end of the M-phase). G1/S cyclins are essential for the control of the cell cycle at the G1/S (start) transition.

j) Eukaryotic thiol (cysteine) proteases active sites (Cys-protease). Some SEQ ID NOS, and thus also the sequences they validate, repreasent polynucleotides encoding

proteins having a eukaryotic thiol (cysteine) protease active site. Eukaryotic thiol proteases (Dufour E., *Biochimie* (1988) *70*:1335); are a family of proteolytic enzymes which contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad. The proteases that belong to this family are: 1) vertebrate lysosomal cathepsins B (Kirschke H., et al., *Protein Prof.* (1995) *2*:1587-1643); 2) vertebrate lysosomal dipeptidyl peptidase I (also known as cathepsin C) (Kirschke H., et al., *supra*); 3) vertebrate calpains (Calpains are intracellular calcium-activated thiol protease that contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain); 4) mammalian cathepsin K, which seems involved in osteoclastic bone resorption (Shi G.-P., et al., *FEBS Lett.* (1995) *357*:129); 5) human cathepsin O ([4] Velasco G., Ferrando A.A., Puente X.S., Sanchez L.M., Lopez-Otin C. *J. Biol. Chem.* (1994) *269*:27136); 6) bleomycin hydrolase (which catalyzes the inactivation of the antitumor drug BLM (a glycopeptide)); 7) Plant enzymes such as: barley aleurain, EP-B1/B4; kidney bean EP-C1, rice bean SH-EP; kiwi fruit actinidin; papaya latex papin, chymopapain, caricain, and proteinase IV; pea turgor-responsive protein 15A; pineapple stem bromelain; rape COT44; rice oryzain alpha, beta, and gamma; tomato low-temperature induced, Arabidopsis thaliana A494, RD19A and RD21A; 8) - House-dust mites allergens DerP1 and EurM1; 9) cathepsin B-like proteinases from the worms Caenorhabditis elegans (genes gcp-1, cpr-3, cpr-4, cpr-5 and cpr-6), Schistosoma mansoni (antigen SM31) and Japonica (antigen SJ31), Haemonchus contortus (genes AC-1 and AC-2), and Ostertagia ostertagi (CP-1 and CP-3); 10) slime mold cysteine proteinases CP1 and CP2; 11) cruzipain from Trypanosoma cruzi and brucei; 12) throphozoite cysteine proteinase (TCP) from various Plasmodium species; 13) proteases from Leishmania mexicana, Theileria annulata and Theileria parva; 14) Baculoviruses cathepsin-like enzyme (v-cath); 15) Drosophila small optic lobes protein (gene sol), a neuronal protein that contains a calpain-like domain; 16) yeast thiol protease BLH1/YCP1/LAP3; 17) Caenorhabditis elegans hypothetical protein C06G4.2, a calpain-like protein.

In addition, two bacterial peptidases are also part of this family: 1) aminopeptidase C from Lactococcus lactis (gene pepC) (Chapot-Chartier M.P., et al., *Appl. Environ. Microbiol.* (1993) *59*:330); and 2) thiol protease tpr from Porphyromonas gingivalis. Three other proteins are structurally related to this family, but may have lost their proteolytic

activity. These include: 1) soybean oil body protein P34 (which has its active site cysteine replaced by a glycine); 2) rat testin (which is a sertoli cell secretory protein highly similar to cathepsin L but with the active site cysteine is replaced by a serine); and 3) Plasmodium falciparum serine-repeat protein (SERA) (which is the major blood stage antigen and possesses a C-terminal thiol-protease-like domain (Higgins D.G., et al., *Nature* (1989) *340*:604), with the active site cysteine is replaced by a serine).

k) Phorbol Esters/Diacylglycerol Binding (DAG_PE_bind). One SEQ represents a polynucleotide encoding a protein belonging to the family including phorbol esters/diacylglycerol binding proteins. Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC) (Azzi *et al.*, *Eur. J. Biochem.* (1992) *208*:547). Phorbol esters can directly stimulate PKC. The N-terminal region of PKC, known as C1, has been shown (Ono *et al.*, *Proc. Natl. Acad. Sci. USA* (1989) *86*:4868) to bind PE and DAG in a phospholipid and zinc-dependent fashion. The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. Such a domain has also been found in, for example, the following proteins.

(1) Diacylglycerol kinase (EC 2.7.1.107) (DGK) (Sakane *et al.*, *Nature* (1990) *344*:345), the enzyme that converts DAG into phosphatidate. It contains two copies of the DAG/PE-binding domain in its N-terminal section. At least five different forms of DGK are known in mammals; and

(2) N-chimaerin, a brain specific protein which shows sequence similarities with the BCR protein at its C-terminal part and contains a single copy of the DAG/PE-binding domain at its N-terminal part. It has been shown (Ahmed *et al.*, *Biochem. J.* (1990) *272*:767, and Ahmed *et al.*, *Biochem. J.* (1991) *280*:233) to be able to bind phorbol esters.

The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in this domain. The signature pattern completely spans the DAG/PE domain.

1) DEAD and DEAH box families ATP-dependent helicases signatures (Dead_box_helic). Some SEQ ID NOS represent polynucleotides encoding a novel member of the DEAD box family. A number of eukaryotic and prokaryotic proteins have been characterized (Schmid S.R., et al., *Mol. Microbiol.* (1992) *6*:283; Linder P., et al., *Nature* (1989) *337*:121; Wassarman D.A., et al., *Nature* (1991) *349*:463) on the basis of their structural similarity. All are involved in ATP-dependent, nucleic-acid unwinding. Proteins currently known to belong to this family are:

1) Initiation factor eIF-4A. Found in eukaryotes, this protein is a subunit of a high molecular weight complex involved in 5'cap recognition and the binding of mRNA to ribosomes. It is an ATP-dependent RNA-helicase.

2) PRP5 and PRP28. These yeast proteins are involved in various ATP-requiring steps of the pre-mRNA splicing process.

3) Pl10, a mouse protein expressed specifically during spermatogenesis.

4) An3, a Xenopus putative RNA helicase, closely related to Pl10.

5) SPP81/DED1 and DBP1, two yeast proteins involved in pre-mRNA splicing and related to Pl10.

6) Caenorhabditis elegans helicase glh-1.

7) MSS116, a yeast protein required for mitochondrial splicing.

8) SPB4, a yeast protein involved in the maturation of 25S ribosomal RNA.

9) p68, a human nuclear antigen. p68 has ATPase and DNA-helicase activities in vitro. It is involved in cell growth and division.

10) Rm62 (p62), a Drosophila putative RNA helicase related to p68.

11) DBP2, a yeast protein related to p68.

12) DHH1, a yeast protein.

13) DRS1, a yeast protein involved in ribosome assembly.

14) MAK5, a yeast protein involved in maintenance of dsRNA killer plasmid.

15) ROK1, a yeast protein.

16) ste13, a fission yeast protein.

17) Vasa, a Drosophila protein important for oocyte formation and specification of embryonic posterior structures.

18) Me31B, a Drosophila maternally expressed protein of unknown function.

19) dbpA, an Escherichia coli putative RNA helicase.

20) deaD, an Escherichia coli putative RNA helicase which can suppress a mutation in the rpsB gene for ribosomal protein S2.

21) rhlB, an Escherichia coli putative RNA helicase.

22) rhlE, an Escherichia coli putative RNA helicase.

23) rmB, an Escherichia coli protein that shows RNA-dependent ATPase activity, which interacts with 23S ribosomal RNA.

24) Caenorhabditis elegans hypothetical proteins T26G10.1, ZK512.2 and ZK686.2.

25) Yeast hypothetical protein YHR065c.

26) Yeast hypothetical protein YHR169w.

27) Fission yeast hypothetical protein SpAC31A2.07c.

28) Bacillus subtilis hypothetical protein yxiN.

All of the above proteins share a number of conserved sequence motifs. Some of them are specific to this family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman T.C., *Nature* (1988) *333*:22 and *Nature* (1988) *333*:578 (Errata); http://www.expasy.ch/www/linder/HELICASES_TEXT.html). One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins (Wassarman D.A., et al., *Nature* (1991) *349*:463; Harosh I., et al., *Nucleic Acids Res.* (1991) *19*:6331; Koonin E.V., et al., *J. Gen. Virol.* (1992) *73*:989). Proteins currently known to belong to this DEAH subfamily are:

1) PRP2, PRP16, PRP22 and PRP43. These yeast proteins are all involved in various ATP-requiring steps of the pre-mRNA splicing process. 2) Fission yeast prh1, which my be involved in pre-mRNA splicing. 3) Male-less (mle), a Drosophila protein required in males, for dosage compensation of X chromosome linked genes. 4) RAD3 from yeast. RAD3 is a DNA helicase involved in excision repair of DNA damaged by UV light, bulky adducts or cross-linking agents. Fission yeast rad15 (rhp3) and mammalian DNA excision repair protein XPD (ERCC-2) are the homologs of RAD3. 5) Yeast CHL1 (or CTF1), which is important for chromosome transmission and normal cell cycle progression in G(2)/M. 6) Yeast TPS1. 7) Yeast hypothetical protein YKL078w. 8) Caenorhabditis elegans hypothetical proteins C06E1.10 and K03H1.2. 9) Poxviruses' early

transcription factor 70 Kd subunit which acts with RNA polymerase to initiate transcription from early gene promoters. 10) I8, a putative vaccinia virus helicase. 11) hrpA, an Escherichia coli putative RNA helicase.

m) EF Hand (EFhand). Several of the validation sequences, and thus the sequences they validate, correspond to polynucleotides encoding a novel protein in the family of EF-hand proteins. Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand (Kawasaki *et al.*, *Protein. Prof.* (1995) *2*:305-490). This type of domain consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand).

Proteins known to contain EF-hand regions include: Calmodulin (Ca=4, except in yeast where Ca=3) ("Ca=" indicates approximate number of EF-hand regions); diacylglycerol kinase (EC 2.7.1.107) (DGK) (Ca=2); 2) FAD-dependent glycerol-3-phosphate dehydrogenase (EC 1.1.99.5) from mammals (Ca=1); guanylate cyclase activating protein (GCAP) (Ca=3); MIF related proteins 8 (MRP-8 or CFAG) and 14 (MRP-14) (Ca=2); myosin regulatory light chains (Ca=1); oncomodulin (Ca=2); osteonectin (basement membrane protein BM-40) (SPARC); and proteins that contain an "osteonectin" domain (QR1, matrix glycoprotein SC1).

n) Ets Domain (Ets_Nterm). One SEQ ID NO, and thus the sequence it validates, represents a polynucleotide encoding a polypeptide with N-terminal homology in ETS domain. Proteins of this family contain a conserved domain, the "ETS-domain," that is involved in DNA binding. The domain appears to recognize purine-rich sequences; it is about 85 to 90 amino acids in length, and is rich in aromatic and positively charged residues (Wasylyk, et al., , *Eur. J. Biochem.* (1993) *211*:718).

The *ets* gene family encodes a novel class of DNA-binding proteins, each of which binds a specific DNA sequence. These proteins comprise an *ets* domain that specifically interacts with sequences containing the common core tri-nucleotide sequence GGA. In addition to an *ets* domain, native *ets* proteins comprise other sequences which can modulate

the biological specificity of the protein. *Ets* genes and proteins are involved in a variety of essential biological processes including cell growth, differentiation and development, and three members are implicated in oncogenic process.

o) Type II fibronectin collagen-binding domain (FntypeII). A few of the validation sequences, and thus the sequences they validate, represent polynucleotides encoding a polypeptide having a type II fibronectin collagen binding domain. Fibronectin is a plasma protein that binds cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. The major part of the sequence of fibronectin consists of the repetition of three types of domains, which are called type I, II, and III (Skorstengaard K., et al., *Eur. J. Biochem.* (1986) *161*:441). Type II domain is approximately forty residues long, contains four conserved cysteines involved in disulfide bonds and is part of the collagen-binding region of fibronectin. In fibronectin the type II domain is duplicated. Type II domains have also been found in the following proteins: 1) blood coagulation factor XII (Hageman factor) (1 copy); 2) bovine seminal plasma proteins PDC-109 (BSP-A1/A2) and BSP-A3 (Seidah N.G., et al., *Biochem. J.* (1987) *243*:195. (twice); 3) cation-independent mannose-6-phosphate receptor (which is also the insulin-like growth factor II receptor) Kornfeld S., *Annu. Rev. Biochem.* (1992) *61*:307) (1 copy); 4) Mannose receptor of macrophages (Taylor M.E., et al., *J. Biol. Chem.* (1990) *265*:12156) (1 copy); 5) 180 Kd secretory phospholipase A2 receptor (1 copy) Lambeau G., et al., *J. Biol. Chem.* (1994) *269*:1575; 6) DEC-205 receptor (1 copy); 6) Jiang W., et al., *Nature* (1995) *375*:151); 7) 72 Kd type IV collagenase (EC 3.4.24.24) (MMP-2) (Collier I.E., et al., *J. Biol. Chem.* (1988) *263*:6579) (3 copies); 7) 92 Kd type IV collagenase (EC 3.4.24.24) (MMP-9) (3 copies); 8) Hepatocyte growth factor activator (Miyazawa K., et al., *J. Biol. Chem.* (1993) *268*:10024) (1 copy).

p) G-Protein Alpha Subunit (G-alpha). Several of the validation sequences, and thus the sequences they validate, correspond to a gene encoding a novel polypeptide of the G-protein alpha subunit family. Guanine nucleotide binding proteins (G-proteins) are a family of membrane-associated proteins that couple extracellularly-activated integral-membrane receptors to intracellular effectors, such as ion channels and enzymes that vary the concentration of second messenger molecules. G-proteins are composed of 3 subunits (alpha, beta and gamma) which, in the resting state, associate as a trimer at the inner face of the plasma membrane. The alpha subunit has a molecule of guanosine diphosphate (GDP)

bound to it. Stimulation of the G-protein by an activated receptor leads to its exchange for GTP (guanosine triphosphate). This results in the separation of the alpha from the beta and gamma subunits, which always remain tightly associated as a dimer. Both the alpha and beta-gamma subunits are then able to interact with effectors, either individually or in a cooperative manner. The intrinsic GTPase activity of the alpha subunit hydrolyses the bound GTP to GDP. This returns the alpha subunit to its inactive conformation and allows it to reassociate with the beta-gamma subunit, thus restoring the system to its resting state.

G-protein alpha subunits are 350-400 amino acids in length and have molecular weights in the range 40-45 kDa. Seventeen distinct types of alpha subunit have been identified in mammals. These fall into 4 main groups on the basis of both sequence similarity and function: alpha-s, alpha-q, alpha-i and alpha-12 (Simon *et al.*, *Science* (1993) *252*:802). Many alpha subunits are substrates for ADP-ribosylation by cholera or pertussis toxins. They are often N-terminally acylated, usually with myristate and/or palmitoylate, and these fatty acid modifications are probably important for membrane association and high- affinity interactions with other proteins. The atomic structure of the alpha subunit of the G-protein involved in mammalian vision, transducin, has been elucidated in both GTP- and GDB-bound forms, and shows considerable similarity in both primary and tertiary structure in the nucleotide-binding regions to other guanine nucleotide binding proteins, such as p21-ras and EF-Tu.

q) <u>Helicases conserved C-terminal domain (helicase_C)</u>. Some SEQ ID NOS, and thus the sequences they validate, represent polynucleotides encoding novel members of the DEAD/H helicase family. The DEAD and DEAH families are described above.

r) <u>Homeobox domain (homeobox).</u> One SEQ ID NO, and thus the sequence it validates, represents a polynucleotide encoding a protein having a homeobox domain. The 'homeobox' is a protein domain of 60 amino acids (Gehring In: <u>Guidebook to the Homebox Genes</u>, Duboule D., Ed., pp1-10, Oxford University Press, Oxford, (1994); Buerglin In: <u>Guidebook to the Homebox Genes</u>, pp25-72, Oxford University Press, Oxford, (1994); Gehring *Trends Biochem. Sci.* (1992) *17*:277-280; Gehring *et alAnnu. Rev. Genet.* (1986) *20*:147-173; Schofield *Trends Neurosci.* (1987) *10*:3-6; http://copan.bioz.unibas.ch/ homeo.html) first identified in number of Drosophila homeotic and segmentation proteins. It is extremely well conserved in many other animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Several proteins that contain a

homeobox domain play an important role in development. Most of these proteins are sequence-specific DNA-binding transcription factors. The homeobox domain is also very similar to a region of the yeast mating type proteins. These are sequence-specific DNA-binding proteins that act as master switches in yeast differentiation by controlling gene expression in a cell type-specific fashion.

A schematic representation of the homeobox domain is shown below. The helix-turn-helix region is shown by the symbols 'H' (for helix), and 't' (for turn).

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxHHHHHHHHtttHHHHHHHHHxxxxxxxxxx
1                                                            60
```

The pattern detects homeobox sequences 24 residues long and spans positions 34 to 57 of the homeobox domain.

x) <u>MAP kinase kinase (mkk).</u> Several validation sequences, and thus the sequences they validate, represent novel members of the MAP kinase kinase family. MAP kinases (MAPK) are involved in signal transduction, and are important in cell cycle and cell growth controls. The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast, invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates. MAPKK regulation studies have led to the discovery of at least four MAPKK convergent pathways in higher organisms. One of these is similar to the yeast pheromone response pathway which includes the ste11 protein kinase. Two other pathways require the activation of either one or both of the serine/threonine kinase-encoded oncogenes c-Raf-1 and c-Mos. Additionally, several studies suggest a possible effect of the cell cycle control regulator cyclin-dependent kinase 1 (cdc2) on MAPKK activity. Finally, MAPKKs are apparently essential transducers through which signals must pass before reaching the nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) *79*:193-207; Nishida *et al.*, *Trends Biochem Sci* (1993) *18*:128-31; Ruderman *Curr Opin Cell Biol* (1993) *5*:207-13; Dhanasekaran *et al.*, *Oncogene* (1998) *17*:1447-55; Kiefer *et al.*, *Biochem Soc Trans* (1997) *25*:491-8; and Hill, *Cell Signal* (1996) *8*:533-44.

y) <u>3'5'-cyclic nucleotide phosphodiesterases signature (PDEase)</u>. One SEQ ID NO, and thus the sequence it validates, represents a polynucleotide encoding a novel 3'5'-cyclic

nucleotide phosphodiesterases (PDEases). PDEases catalyze the hydrolysis of cAMP or cGMP to the corresponding nucleoside 5' monophosphates (Charbonneau H., et al, *Proc. Natl. Acad. Sci. U.S.A.* (1986) *83*:9308). There are at least seven different subfamilies of PDEases (Beavo J.A., et al., *Trends Pharmacol. Sci.* (1990) *11*:150; http://weber.u.washington.edu/~pde/: 1) Type 1, calmodulin/calcium-dependent PDEases; 2) Type 2, cGMP-stimulated PDEases; 3) Type 3, cGMP-inhibited PDEases; 4) Type 4, cAMP-specific PDEases.; 5) Type 5, cGMP-specific PDEases; 6) Type 6, rhodopsin-sensitive cGMP-specific PDEases; and 7) Type 7, High affinity cAMP-specific PDEases.

All PDEase forms share a conserved domain of about 270 residues.

z) Protein Kinase (protkinase). Several validation sequences, and thus the sequences they validate, represent polynucleotides encoding protein kinases. Protein kinases catalyze phosphorylation of proteins in a variety of pathways, and are implicated in cancer. Eukaryotic protein kinases (Hanks S.K., *et al.*, *FASEB J.* (1995) *9*:576; Hunter T., *Meth. Enzymol.* (1991) *200*:3; Hanks S.K., *et al.*, *Meth. Enzymol.* (1991) *200*:38; Hanks S.K., *Curr. Opin. Struct. Biol.* (1991) *1*:369; Hanks S.K., *et al.*, *Science* (1988) *241*:42) are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. Two of the conserved regions are the basis for the signature pattern in the protein kinase profile. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme (Knighton D.R., *et al.*, *Science* (1991) *253*:407). The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases.. A third profile is based on the alignment in (Hanks S.K., *et al.*, *FASEB J.* (1995) *9*:576) and covers the entire catalytic domain.

The protein kinase profile also detects receptor guanylate cyclases and 2-5A-dependent ribonucleases. Sequence similarities between these two families and the eukaryotic protein kinase family have been noticed previously. The profile also detects

Arabidopsis thaliana kinase-like protein TMKL1 which seems to have lost its catalytic activity.

If a protein analyzed includes the two of the above protein kinase signatures, the probability of it being a protein kinase is close to 100%. Eukaryotic-type protein kinases have also been found in prokaryotes such as Myxococcus xanthus (Munoz-Dorado J., *et al.*, *Cell* (1991) *67*:995) and Yersinia pseudotuberculosis. The patterns shown above has been updated since their publication in (Bairoch A., *et al.*, *Nature* (1988) *331*:22).

aa) Ras family proteins (ras). One SEQ ID NO, and thus the sequence it validates, represent polynucleotides encoding the ras family of small GTP/GDP-binding proteins (Valencia et al., 1991, Biochemistry 30:4637-4648). Ras family members generally require a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. Among ras-related proteins, the highest degree of sequence conservation is found in four regions that are directly involved in guanine nucleotide binding. The first two constitute most of the phosphate and Mg2+ binding site (PM site) and are located in the first half of the G-domain. The other two regions are involved in guanosine binding and are located in the C-terminal half of the molecule. Motifs and conserved structural features of the ras-related proteins are described in Valencia et al., 1991, Biochemistry 30:4637-4648.

bb) Thioredoxin family active site (Thioredox). One SEQ ID NO, and thus the sequence it validates, represent a polynucleotide encoding a protein having a thioredoxin family active site. Thioredoxins (Holmgren A., *Annu. Rev. Biochem.* (1985) *54*:237; Gleason F.K., et al., *FEMS Microbiol. Rev.* (1988) *54*:271; Holmgren A. *J. Biol. Chem.* (1989) *264*:13963; Eklund H., et al. *Proteins* (1991) *11*:13) are small proteins of approximately one hundred amino- acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is well conserved.

A number of eukaryotic proteins contain domains evolutionary related to thioredoxin, and all of them are protein disulphide isomerases (PDI). PDI (Freedman R.B., et al., *Biochem. Soc. Trans.* (1988) *16*:96; Kivirikko K.I., et al., *FASEB J.* (1989)

3:1609; Freedman R.B., et al. *Trends Biochem. Sci.* (1994) *19*:331) is an endoplasmic reticulum enzyme that catalyzes the rearrangement of disulfide bonds in various proteins. The various forms of PDI which are currently known are: 1) PDI major isozyme; a multifunctional protein that also function as the beta subunit of prolyl 4-hydroxylase (EC 1.14.11.2), as a component of oligosaccharyl transferase (EC 2.4.1.119), as thyroxine deïodinase, as glutathione-insulin transhydrogenase, and as a thyroid hormone-binding protein; 2) ERp60 (ER-60; 58 Kd microsomal protein), which is a protease; 3) ERp72; and 4) P5.

cc) <u>TNFR/NGFR family cysteine-rich region (TNFR_c6).</u> One SEQ ID NO, and thus the sequence it validates, represent a polynucleotide encoding a protein having a TNFR/NGFR family cysteine-rich region. A number of proteins, some of which are known to be receptors for growth factors, have been found to contain a cysteine-rich domain of about 110 to 160 amino acids in their N-terminal part, that can be subdivided into four (or in some cases, three) modules of about 40 residues containing 6 conserved cysteines. Proteins known to belong to this family (Mallet S., et al., *Immunol. Today* (1991) 12:220; Sprang S.R., *Trends Biochem. Sci.* (1990) 15:366; Krammer P.H., et al., *Curr. Biol.* (1992) 2:383; Bazan J.F., *Curr. Biol.* (1993) *3*:603) are: 1) Tumor Necrosis Factor type I and type II receptors (TNFR) (Both receptors bind TNF-alpha and TNF-beta, but are only similar in the cysteine-rich region.); 2) Shope fibroma virus soluble TNF receptor (protein T2); 3) Lymphotoxin alpha/beta receptor; 4) Low-affinity nerve growth factor receptor (LA-NGFR); 5) CD40 (Bp50), the receptor for the CD40L (or TRAP) cytokine; 6) CD27, the receptor for the CD27L cytokine; 8) CD30, the receptor for the CD30L cytokine; 9) T-cell protein 4-1BB, the receptor for the 4-1BBL putative cytokine; 10) FAS antigen (or APO-1), the receptor for FASL, a protein involved in apoptosis (programmed cell death); 11) T-cell antigen OX40, the receptor for the OX40L cytokine; 12) Wsl-1, a receptor (for a yet undefined ligand) that mediates apoptosis; 13) Vaccinia virus protein A53 (SalF19R).

The six cysteines all involved in intrachain disulfide bonds (Banner D.W., et al, *Cell* (1993) *73*:431). A schematic representation of the structure of the 40 residue module of these receptors is shown below:

xCxxxxxxxxxxxxCxCxxCxxxxxxxxxxCxxxxCxx

where 'C' represents the conserved cysteine involved in a disulfide bond. The signature pattern for the cysteine-rich region is based mainly on the position of the six conserved cysteines in each of the repeats: Consensus pattern: C-x(4,6)-[FYH]-x(5,10)-C-x(0,2)-C-x(2,3)-C-x(7,11)-C-x(4,6)-[DNEQSKP]-x(2)-C (where the six C's are involved in disulfide bonds).

dd) Four Transmembrane Integral Membrane Proteins (transmembrane4). Several of the validation sequences, and thus the sequences they validate, correspond to a sequence encoding a polypeptide that is a member of the 4 transmembrane segments integral membrane protein family (transmembrane 4 family). The transmembrane 4 family of proteins includes a number of evolutionarily-related eukaryotic cell surface antigens (Levy *et al.*, *J. Biol. Chem.*, (1991) *266*:14597; Tomlinson *et al.*, *Eur. J. Immunol.* (1993) *23*:136; Barclay *et al.* The leucocyte antigen factbooks. (1993) Academic Press, London/San Diego). The proteins belonging to this family include: 1) Mammalian antigen CD9 (MIC3), which is involved in platelet activation and aggregation; 2) Mammalian leukocyte antigen CD37, expressed on B lymphocytes; 3) Mammalian leukocyte antigen CD53 (OX-44), which is implicated in growth regulation in hematopoietic cells; 4) Mammalian lysosomal membrane protein CD63 (melanoma-associated antigen ME491; antigen AD1); 5) Mammalian antigen CD81 (cell surface protein TAPA-1), which is implicated in regulation of lymphoma cell growth; 6) Mammalian antigen CD82 (protein R2; antigen C33; Kangai 1 (KAI1)), which associates with CD4 or CD8 and delivers costimulatory signals for the TCR/CD3 pathway; 7) Mammalian antigen CD151 (SFA-1; platelet-endothelial tetraspan antigen 3 (PETA-3)); 8) Mammalian cell surface glycoprotein A15 (TALLA-1; MXS1); 9) Mammalian novel antigen 2 (NAG-2); 10) Human tumor-associated antigen CO-029; 11) Schistosoma mansoni and japonicum 23 Kd surface antigen (SM23 / SJ23).

The members of the 4 transmembrane family share several characteristics. First, they all are apparently type III membrane proteins, which are integral membrane proteins containing an N-terminal membrane-anchoring domain which is not cleaved during biosynthesis and which functions both as a translocation signal and as a membrane anchor. The family members also contain three additional transmembrane regions, at least seven conserved cysteines residues, and are of approximately the same size (218 to 284 residues).

These proteins are collectively know as the "transmembrane 4 superfamily" (TM4) because they span plasma membrane four times.

A schematic diagram of the domain structure of these proteins is as follows:

```
+-+-----+-------+----+-----+-----+--------------------------+-----+----+
| | TMa | Extra | TM2| Cyt | TM3 | Extracellular            | TM4 | Cyt|
+-+-----+-------+----C-----C-----+--------CC---------C-----C---+-----C----+
                 *********
```

where Cyt is the cytoplasmic domain, TMa is the transmembrane anchor; TM2 to TM4 represents transmembrane regions 2 to 4, 'C' are conserved cysteines, and '*'indicates the position of the consensus pattern. The consensus pattern spans a conserved region including two cysteines located in a short cytoplasmic loop between two transmembrane domains:

ee) Trypsin (trypsin). Some SEQ ID NOS, and thus the sequences they validate, correspond to novel serine proteases of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature* (1988) *334*:528). Proteases known to belong to the trypsin family include: 1) Acrosin; 2) Blood coagulation factors VII, IX, X, XI and XII, thrombin, plasminogen, and protein C; 3) Cathepsin G; 4) Chymotrypsins; 5) Complement components C1r, C1s, C2, and complement factors B, D and I; 6) Complement-activating component of RA-reactive factor; 7) Cytotoxic cell proteases (granzymes A to H); 8) Duodenase I; 9) Elastases 1, 2, 3A, 3B (protease E), leukocyte (medullasin).; 10) Enterokinase (EC 3.4.21.9) (enteropeptidase); 11) Hepatocyte growth factor activator; 12) Hepsin; 13) Glandular (tissue) kallikreins (including EGF-binding protein types A, B, and C, NGF-gamma chain, gamma-renin, prostate specific antigen (PSA) and tonin); 14) Plasma kallikrein; 15) Mast cell proteases (MCP) 1 (chymase) to 8; 16) Myeloblastin (proteinase 3) (Wegener's autoantigen); 17) Plasminogen activators (urokinase-type, and tissue-type); 18) Trypsins I, II, III, and IV; 19) Tryptases; 20) Snake venom proteases such as ancrod, batroxobin, cerastobin, flavoxobin, and protein C activator; 21) Collagenase from common cattle grub and collagenolytic protease from Atlantic sand fiddler crab; 22) Apolipoprotein(a); 23) Blood fluke cercarial protease; 24)

Drosophila trypsin like proteases: alpha, easter, snake-locus; 25) Drosophila protease stubble (gene sb); and 26) Major mite fecal allergen Der p III. All the above proteins belong to family S1 in the classification of peptidases (Rawlings N.D., *et al.*, *Meth. Enzymol.* (1994) *244*:19) and originate from eukaryotic species. It should be noted that bacterial proteases that belong to family S2A are similar enough in the regions of the active site residues that they can be picked up by the same patterns.

ff) WD Domain, G-Beta Repeats (WD_domain). A few of the validation sequences, and the sequences they validate, represent novel members of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) *56*:615). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition.

In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat). Such a repetitive segment has been shown to exist in a number of other proteins including: human LIS1, a neuronal protein involved in type-1 lissencephaly; and mammalian coatomer beta' subunit (beta'-COP), a component of a cytosolic protein complex that reversibly associates with Golgi membranes to form vesicles that mediate biosynthetic protein transport.

gg) wnt Family of Developmental Signaling Proteins (Wnt_dev_sign). Several of the validation sequences, and thus the sequences they validate, correspond to novel members of the wnt family of developmental signaling proteins. Wnt-1 (previously known as int-1), the seminal member of this family, (Nusse R., *Trends Genet.* (1988) *4*:291) is a proto-oncogene induced by the integration of the mouse mammary tumor virus. It is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous system (CNS). The sequence of wnt-1 is highly conserved in mammals, fish, and amphibians. Wnt-1 was found to be a member of a large family of related proteins (Nusse R., *et al.*, *Cell* (1992) *69*:1073; McMahon A.P.,

Trends Genet. (1992) *8*:1; Moon R.T., *BioEssays* (1993) *15*:91) that are all thought to be developmental regulators. These proteins are known as wnt-2 (also known as irp), wnt-3, -3A, -4, -5A, -5B, -6, -7A, -7B, -8, -8B, -9 and -10. At least four members of this family are present in Drosophila; one of them, wingless (wg), is implicated in segmentation polarity.

All these proteins share the following features characteristics of secretory proteins: a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters. . All sequences known to belong to this family are detected by the provided consensus pattern.

hh) Protein Tyrosine Phosphatase (Y_phosphatase). Several of the validation sequences, and thus the sequences they validate, represent a polynucleotide encoding a protein tyrosine kinase. Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) (Fischer *et al.*, *Science* (1991) *253*:401; Charbonneau *et al.*, *Annu. Rev. Cell Biol.* (1992) *8*:463; Trowbridge, *J. Biol. Chem.* (1991) *266*:23517; Tonks *et al.*, *Trends Biochem. Sci.* (1989) *14*:497; and Hunter, *Cell* (1989) *58*:1013) catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s).

Soluble PTPases include PTPN3 (H1) and PTPN4 (MEG), enzymes that contain an N-terminal band 4.1-like domain and could act at junctions between the membrane and cytoskeleton; PTPN6 (PTP-1C; HCP; SHP) and PTPN11 (PTP-2C; SH-PTP3; Syp), enzymes that contain two copies of the SH2 domain at its N-terminal extremity.

Dual specificity PTPases include DUSP1 (PTPN10; MAP kinase phosphatase-1; MKP-1) which dephosphorylates MAP kinase on both Thr-183 and Tyr-185; and DUSP2 (PAC-1), a nuclear enzyme that dephosphorylates MAP kinases ERK1 and ERK2 on both Thr and Tyr residues.

Structurally, all known receptor PTPases are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. Some of the receptor PTPases contain fibronectin type III (FN-III) repeats, immunoglobulin-like domains, MAM domains or carbonic anhydrase-like domains

in their extracellular region. The cytoplasmic region generally contains two copies of the PTPAse domain. The first seems to have enzymatic activity, while the second is inactive but seems to affect substrate specificity of the first. In these domains, the catalytic cysteine is generally conserved but some other, presumably important, residues are not.

PTPase domains consist of about 300 amino acids. There are two conserved cysteines and the second one has been shown to be absolutely required for activity. Furthermore, a number of conserved residues in its immediate vicinity have also been shown to be important.

ii)Zinc Finger, C2H2 Type (Zincfing_C2H2). Several of the validation sequences, and thus the sequences they validate, correspond to polynucleotides encoding novel members of the of the C2H2 type zinc finger protein family. Zinc finger domains (Klug *et al.*, *Trends Biochem. Sci.* (1987) *12*:464; Evans *et al.*, *Cell* (1988) *52*:1; Payre *et al.*, *FEBS Lett.* (1988) *234*:245; Miller *et al.*, *EMBO J.* (1985) *4*:1609; and Berg, *Proc. Natl. Acad. Sci. USA* (1988) *85*:99) are nucleic acid-binding protein structures first identified in the Xenopus transcription factor TFIIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino acid residues. Two cysteine or histidine residues are positioned at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides.

Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc-dependent DNA or RNA binding property of some members of this class.

Mammalian proteins having a C2H2 zipper include (number in parenthesis indicates number of zinc finger regions in the protein): basonuclin (6), BCL-6/LAZ-3 (6), erythroid krueppel-like transcription factor (3), transcription factors Sp1 (3), Sp2 (3), Sp3 (3) and Sp(4) 3, transcriptional repressor YY1 (4), Wilms' tumor protein (4), EGR1/Krox24 (3), EGR2/Krox20 (3), EGR3/Pilot (3), EGR4/AT133 (4), Evi-1 (10), GLI1 (5), GLI2 (4+), GLI3 (3+), HIV-EP1/ZNF40 (4), HIV-EP2 (2), KR1 (9+), KR2 (9), KR3 (15+), KR4 (14+), KR5 (11+), HF.12 (6+), REX-1 (4), ZfX (13), ZfY (13), Zfp-35 (18), ZNF7 (15),

ZNF8 (7), ZNF35 (10), ZNF42/MZF-1 (13), ZNF43 (22), ZNF46/Kup (2), ZNF76 (7), ZNF91 (36), ZNF133 (3).

In addition to the conserved zinc ligand residues, it has been shown that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld *et al.*, *J. Biomol. Struct. Dyn.* (1993) *11*:557) The best conserved position is found four residues after the second cysteine; it is generally an aromatic or aliphatic residue. The consensus pattern for C2H2 zinc fingers is: C-x(2,4)-C-x(3)-[LIVMFYWC]-x(8)-H-x(3,5)-H. The two C's and two H's are zinc ligands.

jj) Zinc finger, C3HC4 type (RING finger), signature (Zincfing_C3H4). Some SEQ ID NOS, and thus the sequences they validate, represent polynucleotides encoding a polypeptide having a C3HC4 type zinc finger signature. A number of eukaryotic and viral proteins contain this signature, which is primarily a conserved cysteine-rich domain of 40 to 60 residues (Borden K.L.B., et al., *Curr. Opin. Struct. Biol.* (1996) *6*:395) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D structure of the zinc ligation system is uniqueto the RING domain and is refered to as the "cross-brace" motif.

1) Mammalian V(D)J recombination activating protein (RAG1). RAG1 activates the rearrangement of immunoglobulin and T-cell receptor genes.

2) Mouse rpt-1. Rpt-1 is a trans-acting factor that regulates gene expression directed by the promoter region of the interleukin-2 receptor alpha chainor the LTR promoter region of HIV-1.

3) Human rfp. Rfp is a developmentally regulated protein that may function in male germ cell development. Recombination of the N-terminal section of rfp with a protein tyrosine kinase produces the ret transforming protein.

4) Human 52 Kd Ro/SS-A protein. A protein of unknown function from the Ro/SS-A ribonucleoprotein complex. Sera from patients with systemic lupus erythematosus or primary Sjogren's syndrome often contain antibodies that react with the Ro proteins.

5) Human histocompatibility locus protein RING1.

6) Human PML, a probable transcription factor. Chromosomal translocation ofPML with retinoic receptor alpha creates a fusion protein which is thecause of acute promyelocytic leukemia (APL).

7) Mammalian breast cancer type 1 susceptibility protein (BRCA1) ([E1]

http://bioinformatics.weizmann.ac.il/hotmolecbase/entries/brca1.htm).

8) Mammalian cbl proto-oncogene.

9) Mammalian bmi-1 proto-oncogene.

10) Vertebrate CDK-activating kinase (CAK) assembly factor MAT1, a protein that stabilizes the complex between the CDK7 kinase and cyclin H (MAT1 stands for 'Menage A Trois').

11) Mammalian mel-18 protein. Mel-18 which is expressed in a variety of tumorcells is a transcriptional repressor that recognizes and bind a specific DNA sequence.

12) Mammalian peroxisome assembly factor-1 (PAF-1) (PMP35), which is somewhat involved in the biogenesis of peroxisomes. In humans, defects in PAF-1 are responsible for a form of Zellweger syndrome, an autosomal recessivedisorder associated with peroxisomal deficiencies.

13) Human MAT1 protein, which interacts with the CDK7-cyclin H complex.

14) Human RING1 protein.

15) Xenopus XNF7 protein, a probable transcription factor.

16) Trypanosoma protein ESAG-8 (T-LR), which may be involved in the postranscriptional regulation of genes in VSG expression sites or may interact with adenylate cyclase to regulate its activity.

17) Drosophila proteins Posterior Sex Combs (Psc) and Suppressor two of zeste (Su(z)2). The two proteins belong to the Polycomb group of genes needed to maintain the segment-specific repression of homeotic selector genes.

18) Drosophila protein male-specific msl-2, a DNA-binding protein which is involved in X chromosome dosage compensation (the elevation of transcription of the male single X chromosome).

19) Arabidopsis thaliana protein COP1 which is involved in the regulation ofphotomorphogenesis.

20) Fungal DNA repair proteins RAD5, RAD16, RAD18 and rad8.

21) Herpesviruses trans-acting transcriptional protein ICP0/IE110. This protein which has been characterized in many different herpesviruses is a trans-activator and/or - repressor of the expression of many viral and cellular promoters.

22) Baculoviruses protein CG30.

23) Baculoviruses major immediate early protein (PE-38).

24) Baculoviruses immediate-early regulatory protein IE-N/IE-2.

25) Caenorhabditis elegans hypothetical proteins F54G8.4, R05D3.4 and T02C1.1.

26) Yeast hypothetical proteins YER116c and YKR017c.

The signature pattern for the C3HC4 finger is based on the central region of the domain:

Example 17: Differential Expression of Polynucleotides of the Invention: Description of Libraries and Detection of Differential Expression

The relative expression levels of the polynucleotides of the invention was assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 20 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepared the cDNA library, the "nickname" of the library that is used in the tables below (in quotes), and the approximate number of clones in the library.

Table 20 Description of cDNA Libraries

Library (lib #)	Description	Number of Clones in this Clustering
1	Km12 L4 Human Colon Cell Line, High Metastatic Potential (derived from Km12C) "High Colon"	307133
2	Km12C Human Colon Cell Line, Low Metastatic Potential "Low Colon"	284755
3	MDA-MB-231 Human Breast Cancer Cell Line, High Metastatic Potential; micro-metastases in lung "High Breast"	326937
4	MCF7 Human Breast Cancer Cell, Non Metastatic "Low Breast"	318979
8	MV-522 Human Lung Cancer Cell Line, High Metastatic Potential "High Lung"	223620
9	UCP-3 Human Lung Cancer Cell Line, Low Metastatic Potential "Low Lung"	312503

Library (lib #)	Description	Number of Clones in this Clustering
12	Human microvascular endothelial cells (HMEC) – Untreated PCR (OligodT) cDNA library	41938
13	Human microvascular endothelial cells (HMEC) – **Basic fibroblast growth factor** (bFGF) treated PCR (OligodT) cDNA library	42100
14	Human microvascular endothelial cells (HMEC) – Vascular endothelial growth factor (VEGF) treated PCR (OligodT) cDNA library	42825
15	Normal Colon – UC#2 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	34285
16	Colon Tumor – UC#2 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	35625
17	Liver Metastasis from Colon Tumor of UC#2 Patient PCR (OligodT) cDNA library "High Colon Metastasis Tissue"	36984
18	Normal Colon – UC#3 Patient PCR (OligodT) cDNA library "Normal Colon Tumor Tissue"	36216
19	Colon Tumor – UC#3 Patient PCR (OligodT) cDNA library "High Colon Tumor Tissue"	41388
20	Liver Metastasis from Colon Tumor of UC#3 Patient PCR (OligodT) cDNA library "High Colon Metastasis Tissue"	30956

The KM12L4 and KM12C cell lines are described in Example 14 above. The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, *e.g.*, Chandrasekaran *et al.*, *Cancer Res.* (1979) *39*:870 (MDA-MB-231 and MCF-7); Gastpar *et al.*, *J Med Chem* (1998) *41*:4965 (MDA-MB-231 and MCF-7); Ranson *et al.*, *Br J Cancer* (1998) *77*:1586 (MDA-MB-231 and MCF-7); Kuang *et al.*, *Nucleic Acids Res* (1998) *26*:1116 (MDA-MB-231 and MCF-7); Varki *et al.*, *Int J Cancer* (1987) *40*:46 (UCP-3);

Varki *et al.*, *Tumour Biol.* (1990) *11*:327; (MV-522 and UCP-3); Varki *et al.*, *Anticancer Res.* (1990) *10*:637; (MV-522); Kelner *et al.*, *Anticancer Res* (1995) *15*:867 (MV-522); and Zhang *et al.*, *Anticancer Drugs* (1997) *8*:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-treated HMEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMEC were prepared by incubation with 20ng/ml BEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation.

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters. The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac *et al.*, *Genomics* (1996) *37*(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought together computationally. These groups of clones are termed "clusters". Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the "purity" of each cluster can be controlled. For example, artifacts of clustering may occur in computational clustering just as artifacts can occur in "wet-lab" screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1^{st}), and the determining the number of cDNA clones corresponding to the selected cluster in the second library (Clones in 2^{nd}). Differential expression of the selected cluster in the first library relative to the second library is expressed as a "ratio" of percent expression between the two libraries. In general, the "ratio" is calculated by: 1) calculating the percent expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the "number of clones" corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the "depth" of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5 , where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar, Biostatistical Analysis, Prentice Hall, Inc., USA, "Differences between Proportions," pp 296-298 (1974).

Example 18: Polynucleotides Differentially Expressed in High Metastatic Potential Breast Cancer Cells Versus Low Metastatic Breast Cancer Cells

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential breast cancer tissue and low metastatic breast cancer cells. Expression of these sequences in breast cancer can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these

polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential breast cancer cells and low metastatic potential breast cancer cells.

Table 21. Differentially expressed polynucleotides: Higher expression in high metastatic potential breast cancer (lib3) relative to low metastatic breast cancer cells (lib4)

SEQ ID NOS:	Sequence Name	Cluster ID	Lib3 clones	Lib4 clones	lib3/lib4	Zscore
889	RTA00000197AR.f.12.1	3513	17	5	3.317240	2.287632
990	RTA00000185AF.a.19.2	5749	9	0	8.780930	2.629923
998	RTA00000196F.e.7.1	1039	10	2	4.878294	1.978215
1003	RTA00000182AF.l.12.1	1027	41	17	2.353059	2.926571
1009	RTA00000192AF.g.23.1	6455	6	0	5.853953	2.011224
1018	RTA00000181AF.e.22.3	3442	17	4	4.146550	2.562391
1027	RTA00000198AF.c.17.1	6923	6	0	5.853953	2.011224
1208	RTA00000187AF.g.13.1	2991	10	1	9.756589	2.371428
1210	RTA00000192AF.o.19.1	3549	10	1	9.756589	2.371428
1231	RTA00000191AF.j.14.1	1002	42	20	2.048883	2.570309
1340	RTA00000190AF.p.3.1	2378	34	0	33.17240	5.588184
1354	RTA00000178AF.n.23.1	3298	12	1	11.70790	2.729313
1356	RTA00000191AF.c.3.1	3549	10	1	9.756589	2.371428
1373	RTA00000178AF.b.13.1	3114	9	1	8.780930	2.174815
1404	RTA00000184AF.i.23.3	1577	25	3	8.130490	3.903813
1450	RTA00000179AR.e.01.4	2493	33	9	3.577416	3.469507
1488	RTA00000197F.i.12.1	3605	14	1	13.65922	3.050936
1490	RTA00000186AF.d.24.1	3114	9	1	8.780930	2.174815
1598	RTA00000187AF.l.11.1	4482	14	3	4.553074	2.374769
1719	RTA00000401F.m.02.1	1573	34	7	4.738914	3.982056
1746	RTA00000422F.c.02.1	2902	18	5	3.512372	2.443314
1765	RTA00000418F.m.19.1	8890	6	0	5.853953	2.011224
1786	RTA00000351R.g.11.1	3077	17	4	4.146550	2.562391
1939	RTA00000408F.l.13.1	4423	12	1	11.70790	2.729313
1948	RTA00000404F.m.10.2	779	60	22	2.660887	3.974953

SEQ ID NOS:	Sequence Name	Cluster ID	Lib3 clones	Lib4 clones	lib3/lib4	Zscore
1975	RTA00000400F.k.22.1	2512	7	0	6.829612	2.235371
2014	RTA00000340R.f.05.1	3202	18	3	5.853953	2.998867
2028	RTA00000422F.c.17.1	1360	26	11	2.306102	2.226876
2049	RTA00000118A.a.23.1	3500	12	3	3.902635	2.018050
2198	RTA00000401F.k.14.1	211	121	43	2.745458	5.856098
2968	RTA00000191AF.j.14.1	1002	42	20	2.048883	2.570309
2379	RTA00000405F.l.11.1	2055	29	8	3.536763	3.213373
2595	RTA00000423F.j.03.1	5391	6	0	5.853953	2.011224
2608	RTA00000399F.o.24.1	2272	17	1	16.58620	3.483575
2621	RTA00000401F.j.15.1	3061	14	0	13.65922	3.428594
2639	RTA00000348R.o.12.1	2263	6	0	5.853953	2.011224
2713	RTA00000340F.f.22.1	1720	57	8	6.951569	5.855075
2726	RTA00000401F.g.22.1	1147	28	12	2.276537	2.294031
2734	RTA00000346F.o.16.1	176	170	44	3.769591	8.366611
2759	RTA00000400F.g.02.1	1508	21	5	4.097767	2.879196
2884	RTA00000527F.j.02.2	4896	11	0	10.73224	2.974502
2903	RTA00000528F.i.22.1	2478	17	5	3.317240	2.287632
3067	RTA00000528F.j.11.1	1070	26	6	4.227855	3.289393
3089	RTA00000527F.k.09.1	213	17	4	4.146550	2.562391
3144	RTA00000528F.b.03.1	2078	11	2	5.366124	2.174565
3169	RTA00000525F.d.13.1	349	77	1	75.12573	8.384408
3306	RTA00000528F.g.22.2	920	76	32	2.317189	4.010278
3332	RTA00000528F.h.02.2	1701	18	4	4.390465	2.714073
3336	RTA00000528F.c.11.1	1701	18	4	4.390465	2.714073

Table 22. Differentially expressed polynucleotides: Higher expression in low metastatic breast cancer cells (lib4) relative to high metastatic potential breast cancer (lib3)

SEQ ID NOS:	Sequence Name	Cluster ID	Lib4 Clones	Lib 3 Clones	lib4/lib3	Zscore
859	RTA00000177AR.n.8.1	4188	4	13	3.33108	1.99126
880	RTA00000181AF.p.4.3	40392	1	8	8.19958	2.03713
888	RTA00000199F.f.08.2	12445	0	11	11.2744	3.05623
933	RTA00000177AF.n.8.3	4188	4	13	3.33108	1.99126
1016	RTA00000186AF.p.09.2	6879	3	43	14.6909	5.83444
1047	RTA00000201F.d.09.1	1827	37	157	4.34910	8.71727
1105	RTA00000192AF.a.24.1	13183	0	7	7.17463	2.30057
1263	RTA00000182AF.j.20.1	4769	2	20	10.2494	3.68254
1264	RTA00000181AF.c.11.1	4769	2	20	10.2494	3.68254
1347	RTA00000197AF.k.9.1	3138	1	10	10.2494	2.45316
1396	RTA00000193AF.b.24.1	35	386	1967	5.22298	33.2328
1408	RTA00000200AF.g.18.1	1600	0	23	23.5738	4.64683
1414	RTA00000183AF.a.19.2	3788	0	6	6.14969	2.07158
1434	RTA00000190AF.d.2.1	2444	26	55	2.16815	3.22244
1537	RTA00000198F.m.12.1	4	987	2807	2.91492	30.3819
1551	RTA00000179AF.p.15.1	5622	2	13	6.66216	2.62993
1555	RTA00000198F.i.2.1	8076	0	9	9.22453	2.70385
1570	RTA00000200R.f.10.1	4	987	2807	2.91492	30.3819
1590	RTA00000178AF.i.01.2	4	987	2807	2.91492	30.3819

SEQ ID NOS:	Sequence Name	Cluster ID	Lib4 Clones	Lib 3 Clones	lib4/lib3	Zscore
1600	RTA00000404F.a.02.1	9738	1	13	13.3243	2.98623
1834	RTA00000126A.o.23.1	6268	3	18	6.14969	3.11179
1966	RTA00000401F.o.06.1	2679	4	23	5.89345	3.52846
1986	RTA00000411F.a.15.1	73812	0	12	12.2993	3.21838
2130	RTA00000345F.n.12.1	7337	3	16	5.46639	2.80694
2133	RTA00000126A.g.7.1	1902	13	48	3.78442	4.45002
2279	RTA00000345F.e.11.1	4392	1	8	8.19958	2.03713
2704	RTA00000340F.p.18.1	287	6	173	29.5526	12.5749
2777	RTA00000400F.f.11.1	4088	0	82	84.0457	9.05778
2778	RTA00000341F.o.12.1	2883	9	21	2.39154	2.07600
2823	RTA00000122A.h.24.1	48	412	1020	2.53749	16.5262
2824	RTA00000346F.j.13.1	5337	5	17	3.48482	2.40321
2851	RTA00000400F.g.08.1	1275	15	32	2.18655	2.41857
2867	RTA00000523F.d.19.1	26489	1	8	8.19958	2.03713
3253	RTA00000526F.d.17.1	2757	4	16	4.09979	2.51500
2064	RTA00000528F.d.04.1	2395	12	37	3.16025	3.51521

Example 19: Polynucleotides Differentially Expressed in High Metastatic Potential Lung Cancer Cells Versus Low Metastatic Lung Cancer Cells

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential lung cancer tissue and low metastatic lung cancer cells. Expression of these sequences in lung cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells are associated can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential lung cancer cells and low metastatic potential lung cancer cells:

Table 23 Differentially expressed polynucleotides: Higher expression in high metastatic potential lung cancer cells (lib8) relative to low metastatic lung cancer cells (lib9)

SEQ ID NO:	Sequence Name	Cluster ID	Lib8 clones	Lib9 clones	lib8/lib9	Zscore
854	RTA00000198AF.n.16.1	3721	9	0	12.5772	3.20845
898	RTA00000200F.o.22.1	983	8	1	11.1797	2.53243
909	RTA00000198AF.m.16.1	51	348	66	7.36849	17.4315
1015	RTA00000198R.c.07.1	19181	6	0	8.38484	2.48169
1047	RTA00000201F.d.09.1	1827	45	15	4.19242	5.09891
1096	RTA00000181AF.e.18.3	8	1355	122	15.5211	39.0214
1097	RTA00000181AF.e.17.3	8	1355	122	15.5211	39.0214
1129	RTA00000181AR.j.14.3	5399	12	0	16.7696	3.80239
1263	RTA00000182AF.j.20.1	4769	10	3	4.65824	2.29362
1264	RTA00000181AF.c.11.1	4769	10	3	4.65824	2.29362
1335	RTA00000196F.k.11.1	3	986	392	3.51507	22.4683
1369	RTA00000198AF.c.7.1	19181	6	0	8.38484	2.48169
1370	RTA00000185AF.e.20.1	5865	12	0	16.7696	3.80239
1396	RTA00000193AF.b.24.1	35	868	11	110.273	34.2897
1537	RTA00000198F.m.12.1	4	506	209	3.38335	15.7309
1544	RTA00000183AF.i.18.2	40129	7	0	9.78231	2.74441
1570	RTA00000200R.f.10.1	4	506	209	3.38335	15.7309
1586	RTA00000177AF.m.1.1	14929	23	16	2.00886	2.02420
1590	RTA00000178AF.i.01.2	4	506	209	3.38335	15.7309
1705	RTA00000339F.f.11.1	5832	5	0	6.98736	2.18988
1834	RTA00000126A.o.23.1	6268	5	0	6.98736	2.18988
1932	RTA00000399F.f.11.1	40167	8	0	11.1797	2.98512
2132	RTA00000423F.e.11.1	2566	11	2	7.68610	2.85611
2261	RTA00000339F.o.07.1	2566	11	2	7.68610	2.85611
2288	RTA00000419F.p.03.1	1937	10	3	4.65824	2.29362
2298	RTA00000340F.l.05.1	38935	7	0	9.78231	2.74441
2414	RTA00000403F.a.17.1	13686	8	0	11.1797	2.98512
2441	RTA00000401F.n.23.1	1552	8	1	11.1797	2.53243
2823	RTA00000122A.h.24.1	48	342	155	3.08345	12.2138
2868	RTA00000528F.b.23.1	1605	22	4	7.68610	4.23808
2878	RTA00000528F.m.16.1	4468	6	1	8.38484	1.97787
2970	RTA00000526F.d.01.1	4468	6	1	8.38484	1.97787

Table 24 Differentially expressed polynucleotides: Higher expression in low metastatic lung cancer cells (lib9) relative to high metastatic potntial lung cancer cells

SEQ ID NO:	Sequence Name	Cluster ID	Lib8 clones	Lib9 clones	lib9/lib8	Zscore
1018	RTA00000181AF.e.22.3	3442	5	23	3.291654	2.368262
1098	RTA00000178AF.n.2.1	17083	0	8	5.724617	2.034117
1310	RTA00000177AF.p.20.1	4141	4	27	4.830145	3.070829
1415	RTA00000198AF.b.14.1	801	16	46	2.057284	2.411087
1418	RTA00000192AF.f.3.1	5257	5	25	3.577885	2.596857
1434	RTA00000190AF.d.2.1	2444	12	37	2.206362	2.299984
1766	RTA00000399F.l.14.1	3354	5	20	2.862308	1.998763
2199	RTA00000406F.m.04.1	14959	11	41	2.667151	2.865855
2266	RTA00000405F.h.07.2	4984	3	16	3.816411	2.058861
2851	RTA00000400F.g.08.1	1275	10	42	3.005423	3.147111
2882	RTA00000527F.p.06.1	1292	8	33	2.951755	2.724411
3089	RTA00000527F.k.09.1	213	137	403	2.104945	7.661033

<u>Example 20:</u> <u>Polynucleotides Differentially Expressed in High Metastatic Potential Colon Cancer Cells Versus Low Metastatic Colon Cancer Cells</u>

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential colon cancer tissue and low metastatic colon cancer cells. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells can be indicative of increased expression of genes or regulatory sequences involved in the metastatic process. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment. In another example, sequences that display higher expression in the low metastatic potential cells can be associated with genes or regulatory sequences that inhibit metastasis, and thus the expression of these polynucleotides in a sample may warrant a more positive prognosis than the gross pathology would suggest.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following table summarizes identified polynucleotides with differential expression between high metastatic potential colon cancer cells and low metastatic potential colon cancer cells:

Table 25 Differentially expressed polynucleotides: Higher expression in high metastatic potential colon cancer (lib1) relative to low metastatic colon cancer cells (lib2)

SEQ ID NO:	Sequence Name	Cluster ID	Lib1 clones	Lib2 clones	lib1/lib2	Zscore
1072	RTA00000187AR.h.15.2	6660	7	0	6.489973399	2.169320547
1124	RTA00000193AF.b.18.1	7542	8	0	7.417112456	2.36964728
1199	RTA00000184AR.b.24.1	5777	9	1	8.344251513	2.09555146
1335	RTA00000196F.k.11.1	3	5268	2164	2.257009497	32.96556438
1447	RTA00000183AR.d.11.3	6420	8	0	7.417112456	2.36964728
1524	RTA00000177AF.f.10.1	6420	8	0	7.417112456	2.36964728
1596	RTA00000192AF.o.7.1	5275	11	2	5.099264814	2.083995588
1597	RTA00000192AF.o.17.1	5275	11	2	5.099264814	2.083995588
2085	RTA00000346F.l.13.1	7542	8	0	7.417112456	2.36964728
2108	RTA00000349R.g.10.1	5777	9	1	8.344251513	2.09555146
2245	RTA00000421F.m.14.1	3524	21	6	3.2449867	2.499690198
2286	RTA00000350R.g.10.1	9026	7	0	6.489973399	2.169320547
2358	RTA00000399F.o.06.1	13574	7	0	6.489973399	2.169320547
2695	RTA00000421F.a.06.1	2385	27	4	6.258188635	3.743586088
2759	RTA00000400F.g.02.1	1508	46	17	2.508729213	3.230059264
2868	RTA00000528F.b.23.1	1605	36	11	3.034273278	3.244010467
2910	RTA00000528F.m.12.1	5768	12	0		3.046665462

Table 26 Differentially expressed polynucleotides: Higher expression in low metastatic colon cancer cells (lib2)relative to high metastatic potential colon cancer (lib1)

SEQ ID NOS:	Sequence Name	Cluster ID	Lib1 clones	Lib2 clones	lib2/lib1	Zscore
877	RTA00000178AR.a.20.1	945	9	21	2.51670	2.21703
1094	RTA00000192AF.j.21.1	2289	3	23	8.26916	3.92187
1126	RTA00000193AF.c.15.1	3726	3	14	5.03340	2.58312
1214	RTA00000179AF.c.15.3	2995	4	13	3.50540	2.09770
1231	RTA00000191AF.j.14.1	1002	12	65	5.84234	6.26259
1287	RTA00000197AR.i.17.1	3516	5	17	3.66719	2.52439
1304	RTA00000179AF.c.15.1	2995	4	13	3.50540	2.09770
1389	RTA00000196F.a.2.1	3575	5	14	3.02004	2.00158
1404	RTA00000184AF.i.23.3	1577	12	40	3.59528	4.01991
1547	RTA00000198F.l.09.1	3611	2	13	7.01081	2.73040
1548	RTA00000190AF.o.12.1	3438	5	14	3.02004	2.00158
1939	RTA00000408F.l.13.1	4423	1	8	8.62869	2.11495
1948	RTA00000404F.m.10.2	779	27	54	2.15717	3.23169
2049	RTA00000118A.a.23.1	3500	3	13	4.67387	2.40298
2198	RTA00000401F.k.14.1	211	109	206	2.03843	6.08597

SEQ ID NOS:	Sequence Name	Cluster ID	Lib1 clones	Lib2 clones	lib2/lib1	Zscore
2231	RTA00000191AF.j.14.1	1002	12	65	5.84234	6.26259
2578	RTA00000345F.b.17.1	945	9	21	2.51670	2.21703
2586	RTA00000422F.b.22.1	2368	14	34	2.61942	3.00662
2798	RTA00000401F.j.23.1	570	59	148	2.70560	6.66631
3106	RTA00000527F.o.12.1	688	29	60	2.23155	3.53946
3169	RTA00000525F.d.13.1	349	69	138	2.15717	5.27497

Example 21: Polynucleotides Differentially Expressed in High Metastatic Potential Colon Cancer Patient Tissue Versus Normal Patient Tissue

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high metastatic potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information. For example, sequences that are highly expressed in the high metastatic potential cells are associated can be indicative of increased expression of genes or regulatory sequences involved in the advanced disease state which involves processes such as angiogenesis, dedifferentiation, cell replication, and metastasis. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant more aggressive treatment.

The differential expression of these polynucleotides can be used as a diagnostic marker, a prognostic marker, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential colon cancer cells and normal colon cells:

Table 27 Differentially expressed polynucleotides isolated from samples from two patients (UC#2 and UC#3) : Higher expression in high metastatic potential colon tissue (UC#2:lib17; UC#3:lib20) vs. normal colon tissue (UC#2:lib15; UC#3:lib18)

SEQ ID NO:	Sequence Name	Cluster ID	lib15 clones	lib17 clones	lib17/lib15	Zscore
909	RTA00000198AF.m.16.1	51	1	10	9.27022	2.28830
2624	RTA00000118A.j.24.1	18	4	23	5.33037	3.27028
2743	RTA00000345F.j.09.1	13	14	80	5.29727	6.34580

SEQ ID NO:	Sequence Name	Cluster ID	lib18 clones	lib20 clones	lib20/lib18	Zscore
2743	RTA00000345F.j.09.1	13	12	23	2.24234	2.16077

Table 28 Differentially expressed polynucleotides isolated from samples from two patients (UC#2 and UC#3) : Higher expression in normal colon tissue (UC#2:lib15; UC#3:lib18)vs. high metastatic potential colon tissue (UC#2:lib17; UC#3:lib20).

SEQ ID NO:	Sequence Name	Cluster ID	Lib5 Clones	L1ib7 Clones	lib15/lib17	Z Score: >2.5899%; >1.96
1335	RTA00000196F.k.11.1	3	242	26	10.04	13.78900072
SEQ ID NO:	Sequence Name	Cluster ID	Lib18 clones	Lib20 clones	lib18/lib20	Zscore
1335	RTA00000196F.k.11.1	3	409	46	7.59993	15.3998

Example 22: Polynucleotides Differentially Expressed in High Colon Tumor Potential Patient Tissue Versus Metastasized Colon Cancer Patient Tissue

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high tumor potential colon cancer tissue and cells derived from high metastatic potential colon cancer cells. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the transformation of precancerous tissue to malignant tissue. This information can be useful in the prevention of achieving the advanced malignant state in these tissues, and can be important in risk assessment for a patient.

The following table summarizes identified polynucleotides with differential expression between high tumor potential colon cancer tissue and cells derived from high metastatic potential colon cancer cells:

Table 29 Differentially expressed polynucleotides: High tumor potential colon tissue vs. metastatic colon tissue

SEQ ID NO:	Sequence Name	Cluster ID	L19 clones	L20 clones	lib19/lib20	Zscore
1096	RTA00000181AF.e.18.3	8	14	1	10.4712	2.56699
1097	RTA00000181AF.e.17.3	8	14	1	10.4712	2.56699
1335	RTA00000196F.k.11.1	3	328	46	5.33318	11.8962
1425	RTA00000191AF.p.3.2	17	24	2	8.97535	3.41950
1537	RTA00000198F.m.12.1	4	26	8	2.43082	2.09705
1570	RTA00000200R.f.10.1	4	26	8	2.43082	2.09705
1590	RTA00000178AF.i.01.2	4	26	8	2.43082	2.09705
2624	RTA00000118A.j.24.1	18	80	13	4.60274	5.51440
2743	RTA00000345F.j.09.1	13	148	23	4.81287	7.68618

Example 23: Polynucleotides Differentially Expressed in High Tumor Potential Colon Cancer Patient Tissue Versus Normal Patient Tissue

A number of polynucleotide sequences have been identified that are differentially expressed between cells derived from high tumor potential colon cancer tissue and normal tissue. Expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. For example, sequences that are highly expressed in the potential colon cancer cells are associated with or can be indicative of increased expression of genes or regulatory sequences involved in early tumor progression. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The following tables summarize polynucleotides that are differentially expressed between high metastatic potential colon cancer cells and normal colon cells:

Table 30 Differentially expressed polynucleotides detected in samples from two patients (UC#2 and UC#3): Higher expression in tumor potential colon tissue (UC#2:lib16; UC#3:lib19)vs. normal colon tissue (UC#2:lib15; UC#3:lib18)

SEQ ID NO:	Sequence Name	Cluster ID	Lib15 clones	Lib16 clones	lib16/lib15	Zscore
2743	RTA00000345F.j.09.1	13	14	50	3.43709	4.22436

SEQ ID NO:	Sequence Name	Cluster ID	Lib18 clones	Lib19 clones	lib19/lib18	Zscore
909	RTA00000198AF.m.16.1	51	0	14	12.2505	3.23250
1096	RTA00000181AF.e.18.3	8	1	14	12.2505	2.84687
1097	RTA00000181AF.e.17.3	8	1	14	12.2505	2.84687
1425	RTA00000191AF.p.3.2	17	4	24	5.25021	3.24580
1537	RTA00000198F.m.12.1	4	6	26	3.79182	2.98901
1560	RTA00000200F.p.05.1	3984	0	7	6.12525	2.09621
1570	RTA00000200R.f.10.1	4	6	26	3.79182	2.98901
1590	RTA00000178AF.i.01.2	4	6	26	3.79182	2.98901
2624	RTA00000118A.j.24.1	18	10	80	7.00028	6.65963
2743	RTA00000345F.j.09.1	13	12	148	10.7921	9.86174

Table 31 Differentially expressed polynucleotides: Higher expression in normal colon tissue (UC#2:lib15) vs. tumor potential colon tissue (UC#2:lib16)

SEQ ID NO:	Sequence Name	Cluster ID	Lib15 clones	Lib16 clones	lib15/lib16	Zscore
1335	RTA00000196F.k.11.1	3	242	39	6.44765	12.3988

Example 24: Polynucleotides Differentially Expressed in Growth Factor-Stimulated Human Microvascular Endothelial Cells (HMEC) Relative to Untreated HMEC

A number of polynucleotide sequences have been identified that are differentially expressed between human microvascular endothelial cells (HMEC) that have been treated with growth factors relative to untreated HMEC.

Sequences that are differentially expressed between growth factor-treated HMEC and untreated HMEC can represent sequences encoding gene products involved in angiogenesis, metastasis (cell migration), and other development and oncogenic processes. For example, sequences that are more highly expressed in HMEC treated with growth factors (such as bFGF or VEGF) relative to untreated HMEC can serve as markers of cancer cells of higher metastatic potential. Detection of expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The following table summarizes identified polynucleotides with differential expression between growth factor-treated and untreated HMEC.

Table 32 Differentially expressed polynucleotides: Higher expression in bFGF treated HMEC (lib13) vs. untreated HMEC (lib12)

SEQ ID NO:	Sequence Name	Cluster ID	Lib12 clones	Lib13 clones	lib13/lib12	Zscore
1492	RTA00000199F.i.9.1	7	25	52	2.07199	2.94741

Table 33 Differentially expressed polynucleotides: Higher expression in VEGF treated HMEC (lib14) vs. untreated HMEC (lib12)

SEQ ID NO:	Sequence Name	Cluster ID	Lib12 clones	Lib14 clones	lib14/lib12	Zscore
1492	RTA00000199F.i.9.1	7	25	67	2.62449	4.17666
2743	RTA00000345F.j.09.1	13	22	49	2.18114	2.99887

Example 25: Polynucleotides Differentially Expressed Across Multiple Libraries

A number of polynucleotide sequences have been identified that are differentially expressed between cancerous cells and normal cells across all three tissue types tested (*i.e.*, breast, colon, and lung). Expression of these sequences in a tissue or any origin can be

valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. These polynucleotides can also serve as non-tissue specific markers of, for example, risk of metastasis of a tumor. The following table summarizes identified polynucleotides that were differentially expressed but without tissue type-specificity in the breast, colon, and lung libraries tested.

Table 34 Polynucleotides Differentially Expressed Across Multiple Library Comparisons

SEQ ID NO.	Cluster	Clones in 1st Lib	Clones in 2nd Lib	Ratio	Cell or Tissue Sample and Cancer State Compared (Z Score)
2868	1605	lib1	lib2	lib1/lib2	colon: high met > low met
		36	11	3.0342732	(3.2440104)
		lib8	lib9	lib8/lib9	lung: high met > low met
		22	4	7.6861036	(4.2380835)
909	51	lib8	lib9	lib8/lib9	lung: high met > low met
		348	66	7.3684960	(17.431560)
		lib18	lib19	lib19/lib18	pt #3 colon: tumor > normal
		0	14	12.250507	(3.2325073)
		lib15	lib17	lib17/lib15	pt #2 colon: met > normal
		1	10	9.2702249	(2.2883061)
1018	3442	lib8	lib9	lib9/lib8	lung: low met > high met
		5	23	3.2916548	(2.3682625)
		lib3	lib4	lib3/lib4	breast: high met > low met
		17	4	4.1465504	(2.5623912)
1047	1827	lib8	lib9	lib8/lib9	lung: high met > low met
		45	15	4.1924201	(5.0989192)
		lib3	lib4	lib4/lib3	breast: low met > high met
		37	157	4.3491051	(8.7172773)
3089	213	lib8	lib9	lib9/lib8	lung: low met > high met
		137	403	2.1049458	(7.6610331)
		lib3	lib4	lib3/lib4	breast: high met > low met
		17	4	4.1465504	(2.5623912)
1834	6268	lib8	lib9	lib8/lib9	lung: high met > low met
		5	0	6.9873669	(2.1898837)
		lib3	lib4	lib4/lib3	breast: low met > high met
		3	18	6.1496901	(3.1117967)
1096	8	lib8	lib9	lib8/lib9	lung: high met > low met
		1355	122	15.521118	(39.021411)
		lib19	lib20	lib19/lib20	pt. #3 colon: tumor > met

SEQ ID NO.	Cluster	Clones in 1st Lib	Clones in 2nd Lib	Ratio	Cell or Tissue Sample and Cancer State Compared (Z Score)
		14	1	10.471247	(2.5669948)
		lib18	lib19	lib19/lib18	pt #3 colon: tumor > normal
		1	14	12.250507	(2.8468716)
1097	8	lib8	lib9	lib8/lib9	lung: high met > low met
		1355	122	15.521118	(39.021411)
		lib19	lib20	lib19/lib20	pt. #3 colon: tumor > met
		14	1	10.471247	(2.5669948)
		lib18	lib19	lib19/lib18	pt #3 colon: tumor > normal
		1	14	12.250507	(2.8468716)
3169	349	lib3	lib4	lib3/lib4	breast: high met > low met
		77	1	75.125736	(8.3844087)
		lib1	lib2	lib2/lib1	colon: low met > high met
		69	138	2.1571737	(5.2749799)
1939	4423	lib3	lib4	lib3/lib4	breast: high met > low met
		12	1	11.707907	(2.7293134)
		lib1	lib2	lib2/lib1	colon: low met > high met
		1	8	8.6286948	(2.1149516)
1968	779	lib3	lib4	lib3/lib4	breast: high met > low met
		60	22	2.6608879	(3.9749537)
		lib1	lib2	lib2/lib1	colon: low met > high met
		27	54	2.1571737	(3.2316908)
1231	1002	lib3	lib4	lib3/lib4	breast: high met > low met
		42	20	2.0488837	(2.5703094)
		lib1	lib2	lib2/lib1	colon: low met > high met
		12	65	5.8423454	(6.2625969)
1263	4769	lib8	lib9	lib8/lib9	lung: high met > low met
		10	3	4.6582446	(2.2936274)
		lib3	lib4	lib4/lib3	breast: low met > high met
		2	20	10.249483	(3.6825426)
1264	4769	lib8	lib9	lib8/lib9	lung: high met > low met
		10	3	4.6582446	(2.2936274)
		lib3	lib4	lib4/lib3	breast: low met > high met
		2	20	10.249483	(3.6825426)
2049	3500	lib3	lib4	lib3/lib4	breast: high met > low met
		12	3	3.9026356	(2.0180506)
		lib1	lib2	lib2/lib1	colon: low met > high met
		3	13	4.6738763	(2.4029818)
1335	3	lib1	lib2	lib1/lib2	colon: high met > low met
		5268	2164	2.2570094	(32.965564)
		lib8	lib9	lib8/lib9	lung: high met > low met
		986	392	3.5150733	(22.468331)

SEQ ID NO.	Cluster	Clones in 1st Lib	Clones in 2nd Lib	Ratio	Cell or Tissue Sample and Cancer State Compared (Z Score)
		lib19	lib20	lib19/lib20	pt #3 colon: tumor > met
		328	46	5.3331820	(11.896271)
		lib18	lib20	lib18/lib20	pt #3 colon: normal > met
		409	46	7.5999342	(15.399861)
		lib15	lib17	lib15/lib17	pt#2 colon: normal > met
		242	26	10.04	(13.789000)
		lib15	lib16	lib15/lib16	pt#2 colon: normal > tumor
		242	39	6.44765	12.39883
1396	35	lib8	lib9	lib8/lib9	lung: high met > low met
		868	11	110.27335	(34.289704)
		lib3	lib4	lib4/lib3	breast: low met > high met
		386	1967	5.2229880	(33.232871)
1404	1577	lib3	lib4	lib3/lib4	breast: high met > low met
		25	3	8.1304909	(3.9038139)
		lib1	lib2	lib2/lib1	colon: low met > high met
		12	40	3.5952895	(4.0199130)
1425	17	lib19	lib20	lib19/lib20	pt #3 colon: tumor > met
		24	2	8.9753551	(3.4195074)
		lib18	lib19	lib19/lib18	pt #3 colon: tumor > normal
		4	24	5.2502174	(3.2458055)
1434	2444	lib3	lib4	lib4/lib3	breast: low met > high met
		26	55	2.1681599	(3.2224421)
		lib8	lib9	lib9/lib8	lung: low met > high met
		12	37	2.2063628	(2.2999846)
2198	211	lib3	lib4	lib3/lib4	breast: high met > low met
		121	43	2.7454588	(5.8560985)
		lib1	lib2	lib2/lib1	colon: low met > high met
		109	206	2.0384302	(6.0859794)
2231	1002	lib3	lib4	lib3/lib4	breast: high met > low met
		42	20	2.0488837	(2.5703094)
		lib1	lib2	lib2/lib1	colon: low met > high met
		12	65	5.8423454	(6.2625969)
1492	7	lib12	lib14	lib14/lib12	HMEC: VEGF > untreated
		25	67	2.6244913	(4.1766696)
		lib12	lib13	lib13/lib12	HMEC: bFGF > untreated
		25	52	2.0719962	(2.9474155)
1537	4	lib8	lib9	lib8/lib9	lung: high met > low met
		506	209	3.3833566	(15.730912)
		lib3	lib4	lib4/lib3	breast: low met > high met
		987	2807	2.9149240	(30.381945)
		lib19	lib20	lib19/lib20	pt#3 colon: tumor > met

SEQ ID NO.	Cluster	Clones in 1st Lib	Clones in 2nd Lib	Ratio	Cell or Tissue Sample and Cancer State Compared (Z Score)
		26	8	2.4308253	(2.0970580)
		lib18	lib19	lib19/lib18	pt#3 colon: tumor > normal
		6	26	3.7918237	(2.9890107)
1570	4	lib8	lib9	lib8/lib9	lung: high met > low met
		506	209	3.3833566	(15.730912)
		lib3	lib4	lib4/lib3	breast: low met > high met
		987	2807	2.9149240	(30.381945)
		lib19	lib20	lib19/lib20	pt#3 colon: tumor > met
		26	8	2.4308253	(2.0970580)
		lib18	lib19	lib19/lib18	pt#3 colon: tumor > normal
		6	26	3.7918237	(2.9890107)
1590	4	lib8	lib9	lib8/lib9	lung: high met > low met
		506	209	3.3833566	(15.730912)
		lib3	lib4	lib4/lib3	breast: low met > high met
		987	2807	2.9149240	(30.381945)
		lib19	lib20	lib19/lib20	pt#3 colon: tumor > met
		26	8	2.4308253	(2.0970580)
		lib18	lib19	lib19/lib18	pt#3 colon: tumor > normal
		6	26	3.7918237	(2.9890107)
2624	18	lib19	lib20	lib19/lib20	pt#3 colon: tumor > met
		80	13	4.6027462	(5.5144093)
		lib18	lib19	lib19/lib18	pt#3 colon: tumor > normal
		10	80	7.0002899	(6.6596394)
		lib15	lib17	lib17/lib15	pt#3 colon: met > normal
		4	23	5.3303793	(3.2702852)
2743	13	lib19	lib20	lib19/lib20	pt#3 colon: tumor > met
		148	23	4.8128716	(7.6861840)
		lib18	lib20	lib20/lib18	pt#3 colon: met > normal
		12	23	2.2423439	(2.1607719)
		lib18	lib19	lib19/lib18	pt#3 colon: tumor > normal
		12	148	10.792113	(9.8617485)
		lib15	lib17	lib17/lib15	pt#2 colon: met > normal
		14	80	5.2972714	(6.3458044)
		lib15	lib16	lib16/lib15	pt#2 colon: tumor > normal
		14	50	3.4370927	(4.2243697)
		lib12	lib14	lib14/lib12	HMEC: VEGF > untreated
		22	49	2.1811410	(2.9988774)
2759	1508	lib1	lib2	lib1/lib2	colon: high met > low met
		46	17	2.5087292	(3.2300592)
		lib3	lib4	lib3/lib4	breast: high met > low met
		21	5	4.0977674	(2.8791960)

SEQ ID NO.	Cluster	Clones in 1st Lib	Clones in 2nd Lib	Ratio	Cell or Tissue Sample and Cancer State Compared (Z Score)
2823	48	lib8	lib9	lib8/lib9	lung: high met > low met
		342	155	3.0834574	(12.213852)
		lib3	lib4	lib4/lib3	breast: low met > high met
		412	1020	2.5374934	(16.526285)
2851	1275	lib3	lib4	lib4/lib3	breast: low met > high met
		15	32	2.1865564	(2.4185764)
		lib8	lib9	lib9/lib8	lung: low met > high met
		10	42	3.0054239	3.1471113

high met = high metastatic potential; low met = low metastatic potential;
met = metastasized; tumor = non-metastasized tumor;
pt = patient; #2 = UC#2; #3 = UC#3;
HMEC = human microvascular endothelial cell;
bFGF = bFGF treated; VEGF = VEGF treated

Example 12: Polynucleotides Exhibiting Colon-Specific Expression

The cDNA libraries described herein were also analyzed to identify those polynucleotides that were specifically expressed in colon cells or tissue, *i.e.*, the polynucleotides were identified in libraries prepared from colon cell lines or tissue, but not in libraries of breast or lung origin. The polynucleotides that were expressed in a colon cell line and/or in colon tissue, but were present in the breast or lung cDNA libraries described herein, are shown in Table 35 (inserted before claims).

Table 35. Polynucleotides Specifically Expressed in Colon

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
847	RTA00000197AF.e.24.1	39250	2	0	0	0	0	0	0	0
851	RTA00000197AR.e.12.1	22095	3	0	0	0	0	0	0	0
860	RTA00000196AF.e.16.1	39252	2	0	0	0	0	0	0	0
862	RTA00000196AF.c.17.1	39602	2	0	0	0	0	0	0	0
865	RTA00000131A.g.19.2	36535	2	0	0	0	0	0	0	0
866	RTA00000187AR.o.10.2	8984	4	3	0	0	0	2	0	0
867	RTA00000198R.b.08.1	22636	3	0	0	0	0	0	0	0
870	RTA00000200R.g.09.1	22785	3	0	0	0	0	0	0	0
873	RTA00000200AF.b.19.1	22847	3	0	0	0	0	0	0	0
875	RTA00000200F.m.15.1	22601	3	0	0	0	1	0	0	0
881	RTA00000181AF.n.15.2	86128	1	0	0	0	0	0	0	0
882	RTA00000196R.k.07.1	22443	2	0	0	0	0	0	0	1
884	RTA00000200AR.e.02.1	36059	2	0	0	0	1	1	1	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
892	RTA00000177AR.a.23.5	6995	4	2	0	0	0	0	0	0
893	RTA00000198R.o.05.1	26702	2	0	0	0	0	0	0	0
894	RTA00000201R.a.02.1	35362	2	0	0	0	0	0	0	0
905	RTA00000197AF.h.11.1	22264	3	0	0	0	0	0	0	0
910	RTA00000199F.c.09.2	16824	3	1	0	0	0	0	0	0
919	RTA00000180AR.h.19.2	84182	1	0	0	0	0	0	0	0
922	RTA00000199R.f.09.1	22907	3	0	0	0	0	0	0	0
923	RTA00000199AF.p.4.1	10282	3	3	0	0	0	0	0	0
929	RTA00000200R.o.03.1	22807	3	0	0	0	0	0	0	0
930	RTA00000189AF.l.22.1	33333	1	1	0	0	0	0	0	0
931	RTA00000195AF.d.20.1	37574	2	0	0	0	0	0	0	0
936	RTA00000198AF.j.18.1	22759	3	0	0	0	0	0	0	0
939	RTA00000180AF.g.3.1	9024	5	2	0	0	0	0	0	0
946	RTA00000199R.j.08.1	37844	2	0	0	0	0	0	0	0
947	RTA00000199F.e.10.1	22906	3	0	0	0	0	0	1	0
949	RTA00000179AF.g.12.3	36390	2	0	0	0	0	0	0	0
952	RTA00000183AR.h.23.2	18957	3	0	0	0	0	0	0	0
953	RTA00000197AF.d.12.1	39546	2	0	0	0	0	0	0	0
960	RTA00000181AR.k.24.3	7005	8	2	0	0	0	0	0	0
963	RTA00000181AR.k.24.2	7005	8	2	0	0	0	0	0	0
968	RTA00000199AR.m.06.1	19122	3	0	0	0	0	0	0	0
973	RTA00000134A.d.10.1	18957	3	0	0	0	0	0	0	0
981	RTA00000181AF.m.4.3	13238	4	1	0	0	0	0	0	0
985	RTA00000196AF.c.6.1	23148	3	0	0	0	0	0	0	0
986	RTA00000198AF.k.19.1	75879	1	0	0	0	0	0	0	0
987	RTA00000199R.h.09.1	76020	1	0	0	0	0	0	0	0
988	RTA00000198AF.o.18.1	13018	4	0	0	0	1	0	0	0
992	RTA00000199F.h.17.2	36254	2	0	0	0	0	0	0	0
993	RTA00000181AR.h.06.3	87226	1	0	0	0	0	0	0	0
1010	RTA00000198AF.f.21.1	22676	3	0	0	0	0	0	0	0
1017	RTA00000200AR.b.07.1	17125	4	0	0	0	0	0	0	0
1022	RTA00000200F.o.03.1	22807	3	0	0	0	0	0	0	0
1024	RTA00000199AF.j.12.1	22461	3	0	0	0	0	0	0	0
1029	RTA00000195AF.d.4.1	22766	3	0	0	0	0	0	0	0
1038	RTA00000200R.k.01.1	40049	2	0	0	0	0	0	0	0
1039	RTA00000198AF.c.10.1	77149	1	0	0	0	0	0	0	0
1042	RTA00000197AR.e.07.1	86969	1	0	0	0	0	0	0	0
1043	RTA00000199R.c.09.1	16824	3	1	0	0	0	0	0	0
1050	RTA00000181AF.o.04.2	22205	3	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1051	RTA00000199AF.l.19.1	22460	3	0	0	0	0	0	0	0
1052	RTA00000198AF.h.22.1	22366	2	1	0	0	0	0	0	0
1055	RTA00000199AF.m.15.1	10101	3	0	0	0	0	0	0	0
1056	RTA00000197AF.j.9.1	13236	4	1	0	0	0	0	0	0
1074	RTA00000185AR.b.18.1	12171	3	2	0	0	0	0	0	0
1079	RTA00000201AF.a.02.1	35362	2	0	0	0	0	0	0	0
1080	RTA00000183AR.h.23.1	18957	3	0	0	0	0	0	0	0
1082	RTA00000187AR.k.12.1	78415	1	0	0	0	0	0	0	0
1086	RTA00000198AF.m.17.1	77992	1	0	0	0	0	0	0	0
1087	RTA00000181AF.m.15.3	12081	4	0	0	0	0	0	0	0
1092	RTA00000198R.c.14.1	39814	2	0	0	0	0	0	0	0
1093	RTA00000200R.o.03.2	22807	3	0	0	0	0	0	0	0
1095	RTA00000192AF.n.13.1	8210	2	6	0	0	0	0	0	0
1100	RTA00000184AR.e.15.1	16347	4	0	0	0	0	0	0	0
1104	RTA00000198R.m.17.1	77992	1	0	0	0	0	0	0	0
1114	RTA00000178R.l.08.1	39648	2	0	0	0	0	0	0	0
1122	RTA00000198AF.p.16.1	71877	1	0	0	0	0	0	0	0
1124	RTA00000193AF.b.18.1	7542	8	0	0	2	1	0	1	0
1128	RTA00000199F.d.10.2	22049	3	0	0	0	0	0	0	0
1131	RTA00000200AF.b.07.1	17125	4	0	0	0	0	0	0	0
1132	RTA00000181AR.i.06.3	19119	3	0	0	0	0	0	0	0
1133	RTA00000196F.k.07.1	22443	2	0	0	0	0	0	0	1
1138	RTA00000198AF.k.23.1	8995	2	5	0	0	0	0	0	0
1140	RTA00000196AF.f.20.1	22774	3	0	0	0	0	0	0	0
1144	RTA00000195AF.c.12.1	37582	2	0	0	0	0	0	0	0
1146	RTA00000186AF.d.1.2	40044	2	0	0	1	0	0	0	0
1151	RTA00000200F.n.05.2	18989	3	0	0	0	0	0	0	0
1152	RTA00000178AF.j.20.1	15066	4	0	0	0	0	0	0	0
1154	RTA00000188AF.m.08.1	22155	3	0	0	0	0	0	0	0
1159	RTA00000199R.d.23.1	37477	2	0	0	0	0	0	0	0
1163	RTA00000200F.n.05.1	18989	3	0	0	0	0	0	0	0
1164	RTA00000196AF.m.13.1	16290	4	0	0	0	0	0	0	0
1169	RTA00000182AF.d.18.4	37435	2	0	0	0	0	0	0	0
1172	RTA00000200AF.g.09.1	22785	3	0	0	0	0	0	0	0
1174	RTA00000177AR.m.17.4	14391	3	1	0	0	0	0	0	0
1175	RTA00000197AR.c.20.1	16282	4	0	0	0	0	0	0	0
1181	RTA00000177AR.m.17.3	14391	3	1	0	0	0	0	0	0
1186	RTA00000196AF.d.10.1	22256	3	0	0	0	0	0	0	0
1187	RTA00000201F.a.18.1	16837	2	2	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1188	RTA00000198AF.o.02.1	68756	1	0	0	0	0	0	0	0
1189	RTA00000187AF.h.21.1	39171	2	0	0	0	0	0	0	0
1191	RTA00000199F.b.03.2	38340	2	0	0	0	0	0	0	0
1202	RTA00000198AF.g.7.1	13386	3	2	0	0	0	0	0	0
1206	RTA00000197AR.c.24.1	82498	1	0	0	0	0	0	0	0
1215	RTA00000197F.e.7.1	86969	1	0	0	0	0	0	0	0
1222	RTA00000181AF.k.24.3	7005	8	2	0	0	0	0	0	0
1226	RTA00000200AF.j.6.1	22902	3	0	0	0	0	0	0	0
1228	RTA00000196AF.h.17.1	39215	2	0	0	0	0	0	0	0
1236	RTA00000185AF.b.11.2	9024	5	2	0	0	0	0	0	0
1241	RTA00000198AF.b.22.1	38956	2	0	0	0	0	0	0	0
1243	RTA00000186AF.m.15.2	40122	2	0	0	0	0	0	0	0
1250	RTA00000199F.f.09.2	22907	3	0	0	0	0	0	0	0
1252	RTA00000183AR.l.15.1	39383	2	0	0	0	0	0	0	0
1257	RTA00000200F.a.12.1	16751	4	0	0	0	0	0	0	0
1260	RTA00000199F.a.5.1	22134	3	0	0	0	0	0	0	0
1262	RTA00000187AR.k.01.1	78356	1	0	0	0	0	0	0	0
1268	RTA00000187AR.j.24.1	78356	1	0	0	0	0	0	0	0
1270	RTA00000199AF.o.19.1	36927	2	0	0	0	0	0	0	0
1273	RTA00000196F.i.19.1	39498	2	0	0	0	0	0	0	0
1274	RTA00000198R.k.23.1	8995	2	5	0	0	0	0	0	0
1276	RTA00000198AF.o.05.1	26702	2	0	0	0	0	0	0	0
1277	RTA00000198R.j.18.1	22759	3	0	0	0	0	0	0	0
1279	RTA00000182AR.c.22.1	16283	3	0	0	0	0	0	0	0
1282	RTA00000180AR.g.03.4	9024	5	2	0	0	0	0	0	0
1295	RTA00000200AF.b.20.1	40403	2	0	0	0	0	0	0	0
1299	RTA00000198AF.d.12.1	21142	2	1	0	0	0	0	0	0
1300	RTA00000200AF.b.12.1	22053	3	0	0	0	0	0	0	0
1301	RTA00000191AR.l.7.2	14391	3	1	0	0	0	0	0	0
1305	RTA00000190AF.e.13.1	38961	2	0	0	0	0	0	0	0
1306	RTA00000196AF.n.17.1	12477	4	1	0	0	0	0	0	0
1311	RTA00000195AF.b.19.1	77678	1	0	0	0	0	0	0	0
1319	RTA00000187AR.m.3.3	17055	4	0	0	0	0	0	0	0
1320	RTA00000200R.g.15.1	22898	3	0	0	0	0	0	0	0
1326	RTA00000187AF.j.7.1	78091	1	0	0	0	0	0	0	0
1329	RTA00000196AF.c.14.1	23105	3	0	0	0	0	0	0	0
1330	RTA00000190AR.p.22.2	16368	4	0	0	0	0	0	0	0
1336	RTA00000198AF.b.8.1	22636	3	0	0	0	0	0	0	0
1337	RTA00000177AF.m.17.1	14391	3	1	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1338	RTA00000200AF.k.1.1	40049	2	0	0	0	0	0	0	0
1342	RTA00000190AF.h.12.1	12977	5	0	0	0	0	0	0	0
1343	RTA00000199F.b.22.2	17018	4	0	0	0	0	0	0	0
1352	RTA00000187AF.i.14.2	19406	2	1	0	0	0	0	0	0
1355	RTA00000196AF.g.10.1	12498	3	1	1	0	0	0	0	0
1361	RTA00000184AF.e.14.1	16347	4	0	0	0	0	0	0	0
1366	RTA00000178AR.h.17.2	23824	2	1	0	0	0	0	0	0
1375	RTA00000195F.a.3.1	27179	2	0	0	0	0	0	0	0
1388	RTA00000196F.j.13.1	23170	3	0	0	0	0	0	0	0
1391	RTA00000196AF.g.8.1	39665	2	0	0	0	0	0	0	0
1393	RTA00000198AF.c.16.1	26801	2	0	0	0	0	0	0	0
1397	RTA00000201F.b.22.1	35728	2	0	0	0	0	0	0	1
1403	RTA00000197AF.p.20.1	22795	3	0	0	0	0	0	0	0
1407	RTA00000192AR.o.16.2	9061	5	2	0	0	0	0	0	0
1409	RTA00000191AF.c.10.1	40422	2	0	0	0	0	0	0	0
1412	RTA00000196AF.p.01.2	87143	1	0	0	0	0	0	0	0
1422	RTA00000180AF.g.17.1	16653	3	1	0	0	0	0	0	0
1427	RTA00000190AR.h.12.2	12977	5	0	0	0	0	0	0	0
1429	RTA00000198AF.n.18.1	16715	3	1	0	0	0	0	0	0
1430	RTA00000199R.o.11.1	23172	3	0	0	0	0	0	0	0
1432	RTA00000191AF.b.4.1	14936	3	0	0	0	0	0	0	0
1433	RTA00000192AF.l.1.1	16392	3	0	0	0	0	0	0	0
1437	RTA00000196R.c.14.2	23105	3	0	0	0	0	0	0	0
1439	RTA00000195R.a.06.1	35265	2	0	1	0	0	0	0	0
1446	RTA00000195AF.b.21.1	39055	2	0	0	0	0	0	0	0
1456	RTA00000197AR.e.22.1	78758	1	0	0	0	0	0	0	0
1459	RTA00000197R.p.20.1	22795	3	0	0	0	0	0	0	0
1462	RTA00000192AF.a.14.1	6874	6	3	0	0	1	0	0	0
1467	RTA00000198R.b.24.1	19047	3	0	0	0	0	0	0	0
1471	RTA00000199F.h.15.2	22269	3	0	0	0	0	0	0	0
1472	RTA00000198AF.g.16.1	6602	1	1	0	0	0	0	0	0
1478	RTA00000192AF.j.6.1	11494	4	0	0	0	0	0	0	0
1479	RTA00000181AF.p.7.3	38773	2	0	0	0	0	0	0	0
1481	RTA00000200AF.g.15.1	22898	3	0	0	0	0	0	0	0
1487	RTA00000184AF.c.9.1	16245	4	0	0	0	0	0	0	0
1489	RTA00000177AF.k.9.1	16245	4	0	0	0	0	0	0	0
1493	RTA00000190AR.l.19.2	88204	1	0	0	0	0	0	0	0
1506	RTA00000201R.a.15.1	57347	1	0	0	0	0	0	0	0
1508	RTA00000195R.a.23.1	86432	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1514	RTA00000186AF.p.17.3	38383	2	0	0	0	0	0	0	0
1518	RTA00000197AR.e.24.1	39250	2	0	0	0	0	0	0	0
1527	RTA00000187AR.j.01.1	79028	1	0	0	0	0	0	0	0
1530	RTA00000201F.f.07.1	51116	1	0	0	0	0	0	0	0
1538	RTA00000201R.c.19.1	22357	2	1	0	0	0	0	0	0
1546	RTA00000177AR.b.8.5	17062	3	0	0	0	0	0	0	0
1556	RTA00000201F.b.21.1	9071	3	4	0	0	0	0	0	0
1561	RTA00000200F.o.10.2	36432	2	0	0	0	0	0	0	0
1562	RTA00000196F.l.14.2	23144	3	0	0	0	0	0	0	0
1569	RTA00000197AF.b.1.1	12134	1	1	0	0	0	0	0	0
1577	RTA00000200AF.d.20.1	26600	2	0	0	0	0	0	0	0
1587	RTA00000178AF.k.9.1	16342	3	0	0	0	0	0	0	0
1592	RTA00000198AF.b.24.1	19047	3	0	0	0	0	0	0	0
1601	RTA00000406F.d.16.1	15040	2	2	0	0	0	0	0	0
1604	RTA00000408F.o.12.2	78578	1	0	0	0	0	0	0	0
1605	RTA00000119A.j.15.1	79623	1	0	0	0	0	0	0	0
1606	RTA00000413F.d.12.1	66467	1	0	0	0	0	0	0	0
1607	RTA00000423F.i.12.1	9118	4	3	0	0	0	0	0	0
1610	RTA00000411F.k.05.1	64777	1	0	0	0	0	0	0	0
1613	RTA00000419F.b.09.1	78128	1	0	0	0	0	0	0	0
1616	RTA00000411F.m.15.1	78014	1	0	0	0	0	0	0	0
1618	RTA00000123A.k.23.1	80313	1	0	0	0	0	0	0	0
1621	RTA00000130A.m.15.1	81630	1	0	0	0	0	0	0	0
1622	RTA00000411F.k.20.1	64973	1	0	0	0	0	0	0	0
1624	RTA00000418F.k.05.1	73021	1	0	0	0	0	0	0	0
1625	RTA00000423F.h.18.1	37972	2	0	0	0	0	0	0	0
1627	RTA00000422F.p.06.2	39282	2	0	0	0	0	0	0	0
1628	RTA00000404F.n.16.2	39095	2	0	0	0	0	0	0	0
1629	RTA00000411F.m.24.1	77568	1	0	0	0	0	0	0	0
1630	RTA00000134A.j.10.1	81383	1	0	0	0	0	0	0	0
1631	RTA00000409F.j.02.1	76417	1	0	0	0	0	0	0	0
1632	RTA00000403F.j.15.1	23840	2	1	0	0	0	0	0	0
1633	RTA00000411F.n.11.1	77276	1	0	0	0	0	0	0	0
1634	RTA00000339F.i.13.1	5970	6	4	0	0	0	0	0	0
1636	RTA00000406F.o.15.1	37482	2	0	0	0	0	0	0	0
1637	RTA00000412F.g.04.2	64457	1	0	0	0	0	0	0	0
1639	RTA00000352R.l.06.1	40343	2	0	0	0	0	0	0	0
1640	RTA00000419F.b.12.1	63148	1	0	0	0	0	0	0	0
1641	RTA00000423F.k.17.2	37512	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1643	RTA00000418F.k.14.1	76133	1	0	0	0	0	1	0	0
1644	RTA00000409F.l.12.1	26755	1	0	0	0	0	0	0	0
1645	RTA00000404F.c.20.1	39088	2	0	0	0	0	0	1	0
1646	RTA00000423F.g.09.1	38958	2	0	0	0	0	0	0	0
1648	RTA00000406F.d.12.1	38575	2	0	0	0	0	0	0	0
1649	RTA00000411F.f.02.1	63386	1	0	0	0	0	0	0	0
1650	RTA00000129A.n.21.1	79381	1	0	0	0	0	0	0	0
1651	RTA00000409F.m.12.1	73490	1	0	0	0	0	0	0	0
1652	RTA00000410F.c.04.1	74099	1	0	0	0	0	0	0	0
1654	RTA00000406F.m.09.1	26891	2	0	0	0	0	0	0	0
1655	RTA00000411F.b.06.1	77884	1	0	0	0	0	0	0	0
1656	RTA00000409F.l.21.1	73143	1	0	0	0	0	0	0	0
1662	RTA00000404F.l.20.2	38638	2	0	0	0	0	0	0	0
1663	RTA00000413F.d.18.1	65305	1	0	0	0	0	0	0	0
1664	RTA00000404F.p.04.2	39069	2	0	0	0	0	0	0	0
1665	RTA00000405F.g.19.2	37150	2	0	0	0	0	0	0	0
1666	RTA00000409F.a.22.1	75200	1	0	0	0	0	0	0	0
1668	RTA00000405F.o.18.1	11016	4	2	0	0	0	0	0	0
1673	RTA00000408F.e.22.2	26930	1	0	0	0	0	0	0	0
1675	RTA00000413F.d.16.1	63331	1	0	0	0	0	0	0	0
1678	RTA00000419F.g.08.1	66700	1	0	0	0	0	0	0	0
1679	RTA00000122A.g.16.1	81366	1	0	0	0	0	0	0	0
1680	RTA00000419F.c.16.1	65254	1	0	0	0	0	0	0	0
1681	RTA00000411F.b.03.1	23634	1	2	0	0	0	0	0	0
1686	RTA00000403F.l.20.1	18267	1	0	0	0	0	0	0	0
1689	RTA00000411F.a.02.1	78537	1	0	0	0	0	0	0	0
1691	RTA00000412F.l.04.1	66372	1	0	0	0	0	0	0	0
1693	RTA00000406F.a.23.1	38712	2	0	0	0	0	0	0	0
1695	RTA00000120A.n.19.3	80004	1	0	0	0	0	0	0	0
1696	RTA00000403F.e.01.1	38965	2	0	0	0	0	0	0	0
1697	RTA00000411F.l.03.1	62702	1	0	0	0	0	0	0	0
1700	RTA00000121A.m.2.1	81064	1	0	0	0	0	0	0	0
1702	RTA00000418F.j.12.1	73316	1	0	0	0	0	0	0	0
1706	RTA00000125A.g.16.1	21497	2	1	0	0	0	0	0	0
1707	RTA00000418F.o.18.1	78676	1	0	0	0	0	0	0	0
1709	RTA00000408F.k.14.1	73856	1	0	0	0	0	0	0	0
1715	RTA00000403F.o.15.1	39140	2	0	0	0	0	0	0	0
1716	RTA00000341F.m.13.1	26502	1	0	0	0	0	0	0	0
1717	RTA00000408F.h.03.1	78382	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1718	RTA00000423F.k.05.1	37472	2	0	0	0	0	0	0	0
1720	RTA00000418F.p.19.1	78544	1	0	0	0	0	0	0	0
1721	RTA00000420F.f.06.1	64812	1	0	0	0	0	0	0	0
1722	RTA00000122A.j.18.1	81317	1	0	0	0	0	0	0	0
1723	RTA00000420F.d.05.1	64432	1	0	0	0	0	0	0	0
1724	RTA00000403F.m.18.1	39185	2	0	0	0	0	0	0	0
1726	RTA00000411F.j.05.1	40709	1	1	0	0	0	0	0	0
1727	RTA00000403F.a.04.1	23529	2	1	0	0	0	0	0	0
1729	RTA00000406F.f.12.1	21895	2	1	0	0	0	0	0	0
1730	RTA00000418F.g.22.1	74837	1	0	0	0	0	0	0	0
1732	RTA00000404F.l.20.1	38638	2	0	0	0	0	0	0	0
1733	RTA00000408F.i.08.2	75811	1	0	0	0	0	0	0	0
1734	RTA00000122A.d.5.1	81155	1	0	0	0	0	0	0	0
1738	RTA00000419F.b.19.1	65534	1	0	0	0	0	0	0	0
1740	RTA00000418F.k.19.1	74932	1	0	0	0	0	0	0	0
1744	RTA00000419F.g.12.1	66171	1	0	0	0	0	0	0	0
1745	RTA00000404F.n.11.2	38001	2	0	0	0	0	0	0	0
1748	RTA00000419F.o.24.1	65092	1	0	0	0	0	0	0	0
1749	RTA00000419F.k.19.1	75447	1	0	0	0	0	0	0	0
1751	RTA00000127A.i.20.1	81418	1	0	0	0	0	0	0	0
1752	RTA00000422F.g.22.1	22561	3	0	0	0	0	0	0	0
1754	RTA00000413F.h.13.1	65190	1	0	0	0	0	0	0	0
1757	RTA00000348R.j.16.1	7005	8	2	0	0	0	0	0	0
1760	RTA00000418F.n.22.1	79062	1	0	0	0	0	0	0	0
1761	RTA00000406F.l.08.1	39016	2	0	0	0	0	0	0	0
1764	RTA00000409F.j.07.1	75190	1	0	0	0	0	0	0	0
1767	RTA00000411F.e.22.1	63638	1	0	0	0	0	0	0	0
1768	RTA00000347F.a.17.1	16723	3	1	0	0	0	0	0	0
1770	RTA00000404F.n.20.1	26865	2	0	0	0	0	0	0	0
1773	RTA00000404F.b.02.1	38984	2	0	0	0	0	0	0	0
1775	RTA00000403F.b.10.1	73268	1	0	0	0	0	0	0	0
1776	RTA00000406F.i.12.1	39080	2	0	0	0	0	0	0	0
1777	RTA00000406F.h.08.1	16228	2	2	0	0	0	0	0	0
1778	RTA00000418F.i.19.1	79180	1	0	0	0	0	0	0	0
1780	RTA00000412F.h.21.1	64348	1	0	0	0	0	0	0	0
1782	RTA00000120A.g.18.1	81255	1	0	0	0	0	0	0	0
1784	RTA00000423F.j.05.1	37958	2	0	0	0	0	0	0	0
1785	RTA00000132A.k.6.1	81284	1	0	0	0	0	0	0	0
1787	RTA00000406F.p.04.1	37458	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1788	RTA00000347F.a.13.1	22446	3	0	0	0	0	0	0	0
1789	RTA00000419F.p.23.1	64748	1	0	0	0	0	0	0	0
1790	RTA00000419F.d.17.1	64353	1	0	0	0	0	0	0	0
1793	RTA00000124A.k.5.1	80252	1	0	0	0	0	0	0	0
1794	RTA00000404F.h.22.1	18735	2	1	0	0	0	0	1	0
1796	RTA00000410F.o.05.1	75262	1	0	0	0	0	0	0	0
1797	RTA00000339R.l.14.1	19119	3	0	0	0	0	0	0	0
1798	RTA00000403F.m.13.2	39077	2	0	0	0	0	0	0	0
1801	RTA00000419F.g.22.1	64515	1	0	0	0	0	0	0	0
1802	RTA00000404F.g.21.1	37947	2	0	0	0	0	0	0	0
1804	RTA00000138A.n.4.1	21920	2	1	0	0	0	0	0	0
1805	RTA00000410F.b.15.1	77100	1	0	0	0	0	0	0	0
1807	RTA00000419F.j.23.1	74470	1	0	0	0	0	0	0	0
1808	RTA00000411F.j.02.1	65310	1	0	0	0	0	0	0	0
1809	RTA00000419F.p.24.1	63477	1	0	0	0	0	0	0	0
1810	RTA00000404F.a.19.1	38624	2	0	0	0	0	0	0	0
1817	RTA00000346F.e.13.1	74653	1	0	0	0	0	0	0	0
1818	RTA00000419F.c.18.1	41394	1	1	0	0	0	0	0	0
1822	RTA00000404F.e.22.1	11344	3	3	0	0	0	0	0	0
1825	RTA00000125A.k.10.1	81644	1	0	0	0	0	0	0	0
1826	RTA00000347F.c.06.1	18846	2	1	0	0	0	0	0	0
1827	RTA00000411F.k.19.1	64200	1	0	0	0	0	0	0	0
1828	RTA00000345F.i.09.1	27250	2	0	0	0	0	0	0	0
1829	RTA00000423F.k.01.1	40426	2	0	0	0	0	0	0	0
1830	RTA00000408F.d.06.1	78997	1	0	0	0	0	0	0	0
1831	RTA00000128A.b.20.1	79761	1	0	0	0	0	0	0	0
1833	RTA00000195AF.d.4.1	22766	3	0	0	0	0	0	0	0
1835	RTA00000403F.h.12.1	15205	2	1	0	0	0	0	0	0
1836	RTA00000119A.j.22.1	80336	1	0	0	0	0	0	0	0
1839	RTA00000126A.n.7.2	79557	1	0	0	1	0	0	0	0
1841	RTA00000404F.j.08.1	39066	2	0	0	0	0	0	0	0
1842	RTA00000410F.c.14.1	77809	1	0	0	0	0	0	0	0
1843	RTA00000120A.g.23.1	81189	1	0	0	0	0	0	0	0
1844	RTA00000195AF.d.20.1	37574	2	0	0	0	0	0	0	0
1846	RTA00000412F.j.17.1	64071	1	0	0	0	0	0	0	0
1848	RTA00000119A.j.10.1	79646	1	0	0	0	0	0	0	0
1854	RTA00000419F.o.16.1	62867	1	0	0	0	0	0	0	0
1856	RTA00000411F.c.17.1	77664	1	0	0	0	0	0	0	0
1857	RTA00000406F.k.15.1	38549	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1858	RTA00000406F.a.02.1	37744	2	0	0	0	0	0	0	0
1860	RTA00000341F.b.06.1	17008	4	0	0	0	0	0	0	0
1861	RTA00000409F.n.14.1	78190	1	0	0	0	0	0	0	0
1863	RTA00000345F.j.08.1	16731	3	1	0	0	0	0	0	0
1865	RTA00000419F.g.15.1	32519	1	1	0	0	0	0	0	0
1866	RTA00000423F.a.19.1	21396	1	2	0	0	0	0	0	0
1868	RTA00000422F.e.08.1	39020	2	0	0	0	0	0	0	0
1869	RTA00000411F.d.15.1	74890	1	0	0	0	0	0	0	0
1871	RTA00000411F.l.15.1	66704	1	0	0	0	0	0	0	0
1873	RTA00000405F.e.08.1	37916	2	0	0	0	1	0	0	0
1874	RTA00000353R.j.24.1	23089	3	0	0	0	0	0	0	0
1876	RTA00000418F.o.06.1	75930	1	0	0	0	0	0	0	0
1877	RTA00000404F.c.10.1	23534	2	1	0	0	0	0	0	0
1878	RTA00000418F.i.21.1	78728	1	0	0	0	0	0	0	0
1880	RTA00000411F.l.13.1	43114	1	1	0	0	0	0	0	0
1881	RTA00000407F.a.24.1	37560	2	0	0	0	0	0	0	0
1882	RTA00000346F.n.06.1	12439	4	0	0	0	0	0	0	0
1883	RTA00000412F.l.21.1	65183	1	0	0	0	0	0	0	0
1884	RTA00000413F.i.02.1	65857	1	0	0	0	0	0	0	0
1885	RTA00000404F.i.19.1	38698	2	0	0	0	0	0	0	0
1887	RTA00000403F.a.11.1	73109	1	0	0	0	0	0	0	0
1889	RTA00000411F.k.16.1	64759	1	0	0	0	0	0	1	0
1890	RTA00000405F.c.01.1	19236	2	0	0	0	0	0	0	0
1891	RTA00000423F.i.18.1	14996	4	0	0	0	0	0	0	0
1894	RTA00000406F.a.07.1	26607	2	0	0	0	0	0	0	0
1895	RTA00000347F.d.06.1	39122	2	0	0	0	0	0	0	0
1896	RTA00000419F.b.18.1	67034	1	0	0	0	0	0	0	0
1897	RTA00000406F.h.07.1	38003	2	0	0	0	0	0	0	0
1898	RTA00000405F.l.15.1	19575	2	1	0	0	0	0	0	0
1899	RTA00000406F.g.17.1	37979	2	0	0	0	0	0	0	0
1902	RTA00000130A.h.22.1	80933	1	0	0	0	0	0	0	0
1905	RTA00000404F.d.13.1	39036	2	0	0	0	0	0	0	0
1908	RTA00000340F.n.01.1	39081	2	0	0	0	0	0	0	0
1909	RTA00000419F.d.06.1	65496	1	0	0	0	0	0	0	0
1910	RTA00000419F.n.09.1	66070	1	0	0	0	0	0	0	0
1911	RTA00000399F.i.08.1	38927	2	0	0	0	0	0	0	0
1913	RTA00000423F.g.13.1	38028	2	0	0	0	0	0	0	0
1916	RTA00000195AF.b.21.1	39055	2	0	0	0	0	0	0	0
1917	RTA00000403F.h.05.1	39096	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1919	RTA00000422F.p.07.2	39024	2	0	0	1	0	0	0	0
1922	RTA00000421F.n.19.1	16409	3	1	0	0	0	0	0	0
1924	RTA00000345F.k.21.1	40204	2	0	0	0	0	0	0	0
1926	RTA00000405F.a.11.1	39124	2	0	0	0	0	0	0	0
1928	RTA00000413F.e.16.1	63836	1	0	0	0	0	0	0	0
1930	RTA00000404F.o.18.2	39110	2	0	0	0	0	0	0	0
1931	RTA00000409F.i.24.1	76967	1	0	0	0	0	0	0	0
1935	RTA00000340F.n.13.1	17055	4	0	0	0	0	0	0	0
1936	RTA00000340F.p.04.1	78533	1	0	0	0	0	0	0	0
1937	RTA00000411F.c.05.1	73368	1	0	0	0	0	0	0	0
1941	RTA00000404F.i.02.1	39015	2	0	0	0	0	0	0	0
1943	RTA00000403F.m.15.2	26901	2	0	0	0	0	0	0	0
1944	RTA00000412F.h.23.2	65118	1	0	0	0	0	0	0	0
1945	RTA00000418F.j.08.1	73382	1	0	0	0	0	0	0	0
1946	RTA00000125A.n.4.1	81984	1	0	0	0	0	0	0	0
1947	RTA00000412F.l.19.1	65825	1	0	0	0	0	0	0	0
1949	RTA00000129A.p.3.1	32644	1	1	0	0	0	0	0	0
1950	RTA00000340F.p.20.1	17008	4	0	0	0	0	0	0	0
1951	RTA00000411F.a.10.1	73073	1	0	0	0	0	0	0	0
1952	RTA00000409F.n.17.1	76725	1	0	0	0	0	0	0	0
1953	RTA00000404F.c.03.2	39198	2	0	0	0	0	0	0	0
1954	RTA00000420F.a.19.1	34192	1	1	0	0	0	0	0	0
1958	RTA00000420F.d.12.1	64095	1	0	0	0	0	0	0	0
1959	RTA00000409F.j.19.1	73792	1	0	0	0	0	0	0	0
1960	RTA00000422F.d.16.1	39133	2	0	0	0	0	0	0	0
1961	RTA00000418F.m.16.1	74986	1	0	0	0	0	0	0	0
1962	RTA00000405F.c.11.1	39068	2	0	0	0	0	0	0	0
1963	RTA00000404F.k.22.1	39084	2	0	0	0	0	0	0	0
1964	RTA00000418F.k.07.1	75067	1	0	0	0	0	0	0	0
1965	RTA00000403F.c.10.1	75261	1	0	0	0	0	0	0	0
1968	RTA00000410F.m.05.1	74964	1	0	0	0	0	0	0	0
1969	RTA00000405F.i.20.1	38532	2	0	0	0	0	0	0	0
1971	RTA00000408F.p.24.1	74286	1	0	0	0	0	0	0	0
1972	RTA00000418F.k.18.1	75385	1	0	0	0	0	0	0	0
1973	RTA00000422F.m.04.1	38702	2	0	0	0	0	0	0	0
1977	RTA00000403F.a.07.1	73559	1	0	0	0	0	0	0	0
1979	RTA00000403F.b.19.1	22327	2	1	0	0	0	0	0	0
1980	RTA00000418F.m.23.1	77195	1	0	0	0	0	0	0	0
1982	RTA00000404F.i.18.1	21912	2	1	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
1983	RTA00000422F.i.14.1	39300	2	0	0	0	0	0	0	0
1984	RTA00000418F.m.14.1	75711	1	0	0	1	0	0	0	0
1985	RTA00000406F.o.12.1	37459	2	0	0	0	0	0	0	0
1987	RTA00000411F.a.07.1	74547	1	0	0	0	0	0	0	0
1988	RTA00000411F.c.02.1	72852	1	0	0	0	0	0	0	0
1990	RTA00000130A.h.16.1	80761	1	0	0	0	0	0	0	0
1991	RTA00000410F.p.23.1	73948	1	0	0	0	0	0	0	0
1992	RTA00000418F.m.24.1	77114	1	0	0	0	0	0	0	0
1994	RTA00000408F.j.19.2	73752	1	0	0	0	0	0	0	0
1996	RTA00000118A.d.17.1	81921	1	0	0	0	0	0	0	0
1997	RTA00000407F.b.04.1	63221	1	0	0	0	0	0	0	0
1998	RTA00000411F.e.07.1	65008	1	0	0	0	0	0	0	0
2000	RTA00000132A.c.11.1	87278	1	0	0	0	0	0	0	0
2001	RTA00000420F.e.16.1	63639	1	0	0	0	0	0	0	0
2003	RTA00000404F.b.11.1	39079	2	0	0	0	0	0	0	0
2004	RTA00000418F.k.17.1	75390	1	0	0	0	0	0	0	0
2005	RTA00000129A.k.12.1	79322	1	0	0	0	0	0	0	0
2006	RTA00000340R.m.07.1	78415	1	0	0	0	0	0	0	0
2007	RTA00000405F.d.14.1	35209	2	0	0	0	0	0	1	0
2008	RTA00000406F.f.11.1	38601	2	0	0	0	0	0	0	0
2009	RTA00000120A.h.5.1	80344	1	0	0	0	0	0	0	0
2011	RTA00000411F.g.06.1	66065	1	0	0	0	0	0	0	0
2012	RTA00000408F.d.16.1	76318	1	0	0	0	0	0	0	0
2015	RTA00000404F.c.19.1	39026	2	0	0	0	0	0	0	1
2017	RTA00000410F.a.01.1	73354	1	0	0	0	0	0	0	0
2018	RTA00000408F.h.08.1	74575	1	0	0	0	0	0	0	0
2019	RTA00000422F.b.16.1	17045	4	0	0	0	0	0	0	0
2020	RTA00000419F.f.10.1	66193	1	0	0	0	0	0	0	0
2021	RTA00000418F.l.04.1	74140	1	0	0	0	0	0	0	0
2022	RTA00000410F.a.16.1	73548	1	0	0	0	0	0	0	0
2023	RTA00000138A.e.13.1	79608	1	0	0	0	0	0	0	0
2024	RTA00000130A.b.5.1	79579	1	0	0	0	0	0	0	0
2025	RTA00000408F.j.15.2	74759	1	0	0	0	0	0	0	0
2026	RTA00000410F.m.20.1	74285	1	0	0	0	0	0	0	0
2029	RTA00000419F.e.04.1	62963	1	0	0	0	0	0	0	0
2031	RTA00000418F.g.05.1	73075	1	0	0	0	0	0	0	0
2032	RTA00000419F.n.02.1	65963	1	0	0	0	0	0	0	0
2035	RTA00000119A.m.15.1	80989	1	0	0	0	0	0	0	0
2038	RTA00000413F.g.23.1	40700	1	1	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2039	RTA00000403F.a.18.1	75726	1	0	0	0	0	0	0	0
2040	RTA00000404F.m.20.2	39144	2	0	0	0	0	0	0	0
2043	RTA00000419F.h.04.1	65034	1	0	0	0	0	0	0	0
2044	RTA00000408F.d.12.1	75782	1	0	0	0	0	0	0	0
2045	RTA00000133A.m.19.2	80167	1	0	0	0	0	0	0	0
2050	RTA00000126A.o.22.1	81752	1	0	0	0	0	0	0	0
2051	RTA00000419F.n.13.1	66026	1	0	0	0	0	0	0	0
2052	RTA00000130A.h.13.1	80790	1	0	0	0	0	0	0	0
2056	RTA00000411F.m.19.1	74924	1	0	0	0	0	0	0	0
2058	RTA00000419F.k.06.1	78493	1	0	0	0	0	0	0	0
2060	RTA00000412F.d.16.1	26829	1	0	0	0	0	0	0	0
2061	RTA00000119A.j.23.1	79835	1	0	0	0	0	0	0	0
2063	RTA00000195AF.c.12.1	37582	2	0	0	0	0	0	0	0
2067	RTA00000423F.c.19.1	40472	2	0	0	0	0	0	0	0
2068	RTA00000405F.g.24.1	39076	2	0	0	0	0	0	0	0
2070	RTA00000419F.c.11.1	65504	1	0	0	0	0	0	0	0
2071	RTA00000135A.f.14.2	79969	1	0	0	0	0	0	0	0
2072	RTA00000403F.a.05.1	18808	1	1	0	0	0	0	0	0
2073	RTA00000405F.e.17.1	38662	2	0	0	0	0	0	0	0
2074	RTA00000411F.d.05.1	75812	1	0	0	0	0	0	0	0
2076	RTA00000418F.d.03.1	76824	1	0	0	0	0	0	0	0
2077	RTA00000418F.h.08.1	76401	1	0	0	0	0	0	0	0
2078	RTA00000418F.m.10.1	79110	1	0	0	0	0	0	0	0
2079	RTA00000411F.i.15.1	31612	1	1	0	0	0	0	0	0
2080	RTA00000413F.i.23.1	63073	1	0	0	0	0	0	0	0
2081	RTA00000411F.e.24.1	64781	1	0	0	0	0	0	0	0
2082	RTA00000406F.g.22.1	38590	2	0	0	0	0	0	0	0
2083	RTA00000126A.n.13.2	79735	1	0	0	0	0	0	0	0
2084	RTA00000419F.a.02.1	77993	1	0	0	0	0	0	0	0
2085	RTA00000346F.l.13.1	7542	8	0	0	2	1	0	1	0
2089	RTA00000120A.d.15.1	80533	1	0	0	0	0	0	0	0
2090	RTA00000418F.f.21.1	75157	1	0	0	0	0	0	0	0
2092	RTA00000129A.d.1.2	80058	1	0	0	0	0	0	0	0
2095	RTA00000419F.m.20.1	76720	1	0	0	0	0	0	0	0
2097	RTA00000406F.e.15.1	39074	2	0	0	0	0	0	0	0
2099	RTA00000411F.c.10.1	73117	1	0	0	0	0	0	0	0
2103	RTA00000413F.d.05.1	64788	1	0	0	0	0	0	0	0
2104	RTA00000121A.o.3.1	81437	1	0	0	0	0	0	0	0
2106	RTA00000420F.e.02.1	40259	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2112	RTA00000126A.k.7.2	79866	1	0	0	0	0	0	0	0
2114	RTA00000419F.l.03.1	79060	1	0	0	0	0	0	0	0
2116	RTA00000118A.a.2.1	38067	2	0	0	0	0	0	0	0
2117	RTA00000410F.m.18.1	76365	1	0	0	0	0	0	0	0
2119	RTA00000406F.c.20.1	38578	2	0	0	0	0	0	0	0
2120	RTA00000413F.b.14.1	66591	1	0	0	0	0	0	0	0
2121	RTA00000406F.c.18.1	14368	2	0	0	0	0	0	0	0
2122	RTA00000418F.j.09.1	76352	1	0	0	0	0	0	0	0
2123	RTA00000419F.f.23.1	65002	1	0	0	0	0	0	0	0
2125	RTA00000411F.a.05.1	76699	1	0	0	0	0	0	0	0
2126	RTA00000419F.m.21.1	77947	1	0	0	0	0	0	0	0
2127	RTA00000405F.n.16.1	21503	2	1	1	0	0	0	0	0
2128	RTA00000422F.o.19.2	13084	3	2	0	0	0	0	0	0
2129	RTA00000408F.n.02.2	76993	1	0	0	0	0	0	0	0
2134	RTA00000119A.g.7.1	83580	1	0	0	0	0	0	0	0
2135	RTA00000411F.i.02.1	66975	1	0	0	0	0	0	0	0
2136	RTA00000408F.l.09.1	75487	1	0	0	0	0	0	0	0
2137	RTA00000423F.g.04.1	23012	2	1	0	0	0	0	0	0
2139	RTA00000418F.i.18.1	78024	1	0	0	0	0	0	0	0
2140	RTA00000411F.h.15.1	65160	1	0	0	0	0	0	0	0
2141	RTA00000410F.i.19.1	78988	1	0	0	0	0	0	0	0
2142	RTA00000419F.k.24.1	75596	1	0	0	0	0	0	0	0
2145	RTA00000409F.i.09.1	75279	1	0	0	0	0	0	0	0
2146	RTA00000419F.h.02.1	63985	1	0	0	0	0	0	0	0
2147	RTA00000413F.b.12.1	64932	1	0	0	0	0	0	0	0
2148	RTA00000121A.h.18.1	16376	4	0	0	0	0	0	0	0
2149	RTA00000411F.n.20.1	75816	1	0	0	0	0	0	0	0
2151	RTA00000411F.n.12.1	73308	1	0	0	0	0	0	0	0
2152	RTA00000408F.j.12.2	18226	1	0	0	0	0	0	0	0
2153	RTA00000409F.i.03.1	75968	1	0	0	0	0	0	0	0
2156	RTA00000409F.j.05.1	74128	1	0	0	0	0	0	0	0
2157	RTA00000419F.m.04.1	74367	1	0	0	0	0	0	0	0
2158	RTA00000418F.k.03.1	78901	1	0	0	0	0	0	0	0
2159	RTA00000419F.d.16.1	64357	1	0	0	0	0	0	0	0
2160	RTA00000420F.e.10.1	65899	1	0	0	0	0	0	0	0
2163	RTA00000418F.k.08.1	18259	1	0	0	0	0	0	0	0
2166	RTA00000410F.c.02.1	75055	1	0	0	0	0	0	0	0
2168	RTA00000403F.h.18.1	39241	2	0	0	0	0	0	0	0
2169	RTA00000405F.n.13.1	23810	2	1	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2170	RTA00000355R.e.14.1	16837	2	2	0	0	0	0	0	0
2171	RTA00000422F.l.03.1	39147	2	0	0	0	0	0	0	0
2173	RTA00000403F.o.14.1	38971	2	0	0	0	0	0	0	0
2177	RTA00000127A.f.11.1	81463	1	0	0	0	0	0	0	0
2179	RTA00000403F.o.07.1	39037	2	0	0	0	0	0	0	0
2180	RTA00000403F.d.19.1	39243	2	0	0	0	0	0	0	0
2182	RTA00000406F.i.17.1	37902	2	0	0	0	0	0	0	0
2183	RTA00000418F.d.22.1	75324	1	0	0	0	0	0	0	0
2184	RTA00000340R.o.12.1	53732	1	0	0	0	0	0	0	0
2185	RTA00000125A.g.24.1	80397	1	0	0	0	0	0	0	0
2186	RTA00000130A.o.21.1	80218	1	0	0	0	0	0	0	0
2187	RTA00000420F.a.23.1	42158	1	1	0	0	0	0	0	0
2188	RTA00000411F.m.18.1	75629	1	0	0	0	0	0	0	0
2189	RTA00000407F.b.22.1	37487	2	0	0	0	0	0	0	0
2190	RTA00000409F.a.16.1	73990	1	0	0	0	0	0	0	0
2192	RTA00000341F.k.12.1	62985	1	0	0	0	0	0	0	0
2193	RTA00000129A.c.18.2	37216	2	0	0	0	0	0	0	0
2194	RTA00000410F.d.10.1	77561	1	0	0	0	0	0	0	0
2195	RTA00000351R.i.03.1	6874	6	3	0	0	1	0	0	0
2196	RTA00000135A.l.1.2	39426	2	0	0	0	0	0	0	0
2197	RTA00000420F.b.18.1	66136	1	0	0	0	0	0	0	0
2200	RTA00000403F.o.13.1	39049	2	0	0	0	0	0	0	0
2201	RTA00000411F.f.06.1	64186	1	0	0	0	0	0	0	0
2203	RTA00000351R.c.13.1	11476	6	0	0	0	0	0	0	0
2206	RTA00000420F.d.16.1	64485	1	0	0	0	0	0	0	0
2207	RTA00000404F.i.12.1	39001	2	0	0	0	0	0	0	0
2208	RTA00000404F.o.10.2	16785	2	2	0	0	0	0	0	0
2209	RTA00000419F.d.07.1	21421	1	2	0	0	0	0	0	0
2210	RTA00000404F.p.02.2	39097	2	0	1	0	0	0	0	0
2211	RTA00000125A.k.14.1	79457	1	0	0	0	0	0	0	0
2212	RTA00000122A.j.22.1	81151	1	0	0	0	0	0	0	0
2213	RTA00000406F.i.13.1	37904	2	0	0	0	0	0	0	0
2214	RTA00000135A.b.23.1	35241	2	0	0	0	0	0	0	0
2217	RTA00000423F.l.04.1	14320	2	0	0	0	0	0	0	0
2218	RTA00000420F.b.04.1	63820	1	0	0	0	0	0	0	0
2220	RTA00000408F.i.18.2	74410	1	0	0	0	0	0	0	0
2222	RTA00000341F.j.05.1	36177	2	0	0	0	0	0	0	0
2223	RTA00000420F.a.16.1	63345	1	0	0	0	0	0	0	0
2225	RTA00000410F.j.01.1	73399	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2226	RTA00000408F.p.21.1	77930	1	0	0	0	0	0	0	0
2227	RTA00000412F.d.19.1	75743	1	0	0	0	0	0	0	0
2228	RTA00000352R.c.04.1	71976	1	0	0	0	0	0	0	0
2229	RTA00000413F.f.19.1	65189	1	0	0	0	0	0	0	0
2230	RTA00000411F.e.03.1	73648	1	0	0	0	0	0	0	0
2233	RTA00000418F.c.04.1	41587	1	1	0	0	0	0	0	0
2234	RTA00000418F.o.17.1	79069	1	0	0	0	0	0	0	0
2235	RTA00000418F.e.21.1	74773	1	0	0	0	0	0	0	0
2236	RTA00000419F.d.14.1	64945	1	0	0	0	0	0	0	0
2240	RTA00000410F.j.20.1	73601	1	0	0	0	0	0	0	0
2243	RTA00000119A.j.9.1	82060	1	0	0	0	0	0	0	0
2247	RTA00000340F.i.13.1	79299	1	0	0	0	0	0	0	0
2248	RTA00000412F.g.03.1	64740	1	0	0	0	0	0	0	0
2249	RTA00000122A.g.17.1	32655	1	1	0	0	0	0	0	0
2251	RTA00000419F.n.12.1	66086	1	0	0	0	0	0	0	0
2254	RTA00000351R.p.14.1	13166	2	3	0	0	0	0	0	0
2255	RTA00000403F.e.08.1	19126	3	0	0	0	0	0	0	0
2256	RTA00000124A.k.20.1	80913	1	0	0	0	0	0	0	0
2257	RTA00000121A.n.2.1	33585	1	1	0	0	0	0	0	0
2258	RTA00000422F.m.24.1	39159	2	0	1	0	1	1	2	2
2259	RTA00000408F.e.24.2	75002	1	0	0	0	0	0	0	0
2262	RTA00000403F.b.12.1	78775	1	0	0	0	0	0	0	0
2263	RTA00000404F.a.09.1	38985	2	0	0	0	0	0	0	0
2265	RTA00000403F.o.19.1	78615	1	0	0	0	0	0	0	0
2268	RTA00000410F.b.10.1	74504	1	0	0	0	0	0	0	0
2270	RTA00000413F.h.12.1	66929	1	0	0	0	0	0	0	0
2271	RTA00000406F.k.14.1	38651	2	0	0	0	0	0	0	0
2273	RTA00000411F.f.17.1	65661	1	0	0	0	0	0	0	0
2274	RTA00000411F.k.10.1	64506	1	0	0	0	0	0	0	0
2275	RTA00000411F.g.21.1	64500	1	0	0	0	0	0	0	0
2276	RTA00000119A.h.24.1	82266	1	0	0	0	0	0	0	0
2278	RTA00000408F.m.22.2	72949	1	0	0	0	0	0	0	0
2281	RTA00000410F.i.17.1	78147	1	0	0	0	0	0	0	0
2284	RTA00000129A.a.13.2	79780	1	0	0	0	0	0	0	0
2285	RTA00000129A.k.21.1	82067	1	0	0	0	0	0	0	0
2286	RTA00000350R.g.10.1	9026	7	0	0	1	0	0	0	0
2287	RTA00000413F.d.23.1	66030	1	0	0	0	0	0	0	0
2291	RTA00000411F.d.10.1	76445	1	0	0	0	0	0	0	0
2292	RTA00000404F.b.19.1	39281	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2293	RTA00000418F.c.07.1	73245	1	0	0	0	0	0	0	0
2294	RTA00000418F.j.15.1	74855	1	0	0	0	0	1	0	0
2297	RTA00000413F.b.16.1	65126	1	0	0	0	0	0	0	0
2299	RTA00000350R.m.14.1	39171	2	0	0	0	0	0	0	0
2300	RTA00000418F.l.11.1	77158	1	0	0	0	0	0	0	0
2301	RTA00000130A.d.5.1	82051	1	0	0	0	0	0	0	0
2302	RTA00000339F.n.05.1	39648	2	0	0	0	0	0	0	0
2304	RTA00000407F.a.23.1	23489	2	1	0	0	0	0	0	0
2306	RTA00000403F.h.11.1	39219	2	0	0	0	0	0	0	0
2307	RTA00000406F.j.13.1	38688	2	0	0	0	0	0	0	0
2308	RTA00000352R.p.09.1	16915	4	0	0	0	0	0	0	0
2309	RTA00000413F.g.24.1	65481	1	0	0	0	0	0	0	0
2313	RTA00000420F.a.08.1	19473	1	2	0	0	0	0	0	0
2316	RTA00000404F.i.22.1	39082	2	0	0	0	0	0	0	0
2317	RTA00000124A.k.23.1	81350	1	0	0	0	0	0	0	0
2318	RTA00000404F.e.11.1	38991	2	0	0	0	0	0	0	0
2319	RTA00000129A.d.2.4	80119	1	0	0	0	0	0	0	0
2322	RTA00000419F.o.15.1	32487	1	1	0	0	0	0	0	0
2323	RTA00000119A.m.17.1	79536	1	0	0	0	0	0	0	0
2324	RTA00000410F.b.07.1	78916	1	0	0	0	0	0	0	0
2325	RTA00000420F.b.19.1	36873	2	0	0	0	0	0	0	0
2327	RTA00000411F.b.21.1	10051	1	0	0	0	0	0	0	0
2329	RTA00000356R.c.16.1	16915	4	0	0	0	0	0	0	0
2331	RTA00000412F.h.11.1	63175	1	0	0	0	0	0	0	0
2334	RTA00000420F.a.11.1	66460	1	0	0	0	0	0	0	0
2335	RTA00000120A.c.7.1	80985	1	0	0	1	0	0	0	0
2336	RTA00000404F.e.15.1	39101	2	0	0	0	0	0	0	0
2337	RTA00000422F.n.20.1	38676	2	0	0	0	0	0	1	0
2338	RTA00000423F.h.20.1	38639	2	0	0	0	0	0	0	0
2341	RTA00000410F.b.18.1	76701	1	0	0	0	0	0	0	0
2343	RTA00000423F.g.15.1	35173	2	0	0	0	0	0	0	0
2344	RTA00000413F.b.04.1	66427	1	0	0	0	0	0	0	0
2347	RTA00000346F.f.11.1	38528	2	0	0	0	0	0	0	0
2350	RTA00000422F.i.02.1	76436	1	0	0	0	0	0	0	0
2351	RTA00000410F.a.08.1	73324	1	0	0	0	0	0	0	0
2353	RTA00000419F.e.02.1	65010	1	0	0	0	0	0	0	0
2355	RTA00000403F.g.13.1	38718	2	0	0	0	0	0	0	0
2357	RTA00000407F.a.01.1	12501	3	1	0	0	0	0	0	0
2360	RTA00000411F.f.14.1	62984	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2361	RTA00000411F.c.04.1	76858	1	0	0	0	0	0	0	0
2362	RTA00000135A.m.18.1	19255	2	0	0	0	0	0	0	0
2363	RTA00000413F.c.17.1	36831	2	0	0	0	0	0	0	0
2365	RTA00000404F.j.01.1	26859	2	0	0	0	0	0	0	0
2366	RTA00000138A.p.10.1	81625	1	0	0	0	0	0	0	0
2370	RTA00000423F.h.07.1	37933	2	0	0	0	0	0	0	0
2371	RTA00000413F.e.04.1	64176	1	0	0	0	0	0	0	0
2372	RTA00000406F.h.03.1	38585	2	0	0	0	0	0	0	0
2373	RTA00000403F.e.24.1	16432	2	2	0	0	0	0	0	0
2375	RTA00000403F.i.11.1	23535	2	1	0	0	0	0	0	0
2376	RTA00000419F.g.02.1	62839	1	0	0	0	0	0	0	0
2377	RTA00000347F.e.05.1	39814	2	0	0	0	0	0	0	0
2378	RTA00000408F.l.16.1	73468	1	0	0	0	0	0	0	0
2380	RTA00000423F.f.09.1	64823	1	0	0	0	0	0	0	0
2381	RTA00000419F.k.03.1	40822	1	1	0	0	0	0	0	0
2382	RTA00000406F.b.02.1	38744	2	0	0	0	0	0	0	0
2383	RTA00000418F.o.14.1	33524	1	1	0	0	0	0	0	0
2385	RTA00000404F.b.09.1	39166	2	0	0	0	0	0	0	0
2391	RTA00000406F.k.11.1	38715	2	0	0	0	0	0	0	0
2393	RTA00000406F.c.06.1	37924	2	0	0	0	0	0	0	0
2394	RTA00000418F.n.07.1	76316	1	0	0	0	0	0	0	0
2395	RTA00000419F.n.15.1	63484	1	0	0	0	0	0	0	0
2396	RTA00000408F.n.06.2	76642	1	0	0	0	0	0	0	0
2397	RTA00000420F.c.04.1	65007	1	0	0	0	0	0	0	0
2398	RTA00000411F.j.15.1	66871	1	0	0	0	0	0	0	0
2400	RTA00000128A.m.23.1	81441	1	0	0	0	0	0	0	0
2401	RTA00000406F.g.03.1	38690	2	0	0	0	0	0	0	0
2402	RTA00000405F.h.05.2	75706	1	0	0	0	0	0	0	0
2403	RTA00000129A.n.24.1	81409	1	0	0	0	0	0	0	0
2406	RTA00000418F.n.11.1	78977	1	0	0	0	0	0	0	0
2409	RTA00000120A.h.9.1	80736	1	0	0	0	0	0	0	0
2410	RTA00000413F.a.12.1	63403	1	0	0	0	0	0	0	0
2411	RTA00000412F.o.05.1	63575	1	0	0	0	0	0	0	0
2415	RTA00000354R.n.04.1	22049	3	0	0	0	0	0	0	0
2417	RTA00000406F.h.05.1	38542	2	0	0	0	0	0	0	0
2418	RTA00000410F.b.24.1	75104	1	0	0	0	0	0	0	0
2419	RTA00000423F.d.11.1	38950	2	0	0	0	0	0	0	0
2422	RTA00000119A.k.1.1	81282	1	0	0	0	0	0	0	0
2423	RTA00000420F.f.07.1	66312	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2424	RTA00000404F.k.22.2	39084	2	0	0	0	0	0	0	0
2425	RTA00000422F.e.07.1	38964	2	0	0	0	0	0	0	0
2426	RTA00000410F.f.12.1	73883	1	0	0	0	0	0	0	0
2428	RTA00000411F.m.11.1	73196	1	0	0	0	0	0	0	0
2431	RTA00000403F.o.10.2	38964	2	0	0	0	0	0	0	0
2434	RTA00000413F.c.10.1	65600	1	0	0	0	0	0	0	0
2435	RTA00000411F.b.17.1	72893	1	0	0	0	0	0	0	0
2437	RTA00000408F.k.19.1	77593	1	0	0	0	0	0	0	0
2440	RTA00000119A.i.8.1	82593	1	0	0	0	0	0	0	0
2442	RTA00000418F.g.03.1	78737	1	0	0	0	0	0	0	0
2443	RTA00000411F.a.09.1	78629	1	0	0	0	0	0	0	0
2445	RTA00000419F.j.11.1	73183	1	0	0	0	0	0	0	0
2447	RTA00000404F.n.18.2	37169	2	0	0	0	0	0	0	0
2448	RTA00000122A.n.16.1	80553	1	0	0	0	0	0	0	0
2449	RTA00000420F.c.07.1	65555	1	0	0	0	0	0	0	0
2452	RTA00000408F.j.13.2	42275	1	1	0	0	0	0	0	0
2454	RTA00000423F.a.01.1	39103	2	0	0	0	0	0	0	0
2457	RTA00000341F.e.20.1	67422	1	0	0	0	0	0	0	0
2458	RTA00000419F.m.22.1	75600	1	0	0	0	0	0	0	0
2459	RTA00000419F.m.23.1	64263	1	0	0	0	0	0	0	0
2460	RTA00000419F.b.06.1	76728	1	0	0	0	0	0	0	0
2462	RTA00000406F.p.08.1	37573	2	0	0	0	0	0	0	2
2463	RTA00000129A.n.17.1	79811	1	0	0	0	0	0	0	0
2465	RTA00000407F.b.08.1	37513	2	0	0	0	0	0	0	0
2467	RTA00000406F.i.08.1	37946	2	0	0	0	0	0	0	0
2468	RTA00000403F.h.07.1	26856	2	0	0	0	0	0	0	0
2469	RTA00000418F.n.24.1	73153	1	0	0	0	0	0	0	0
2471	RTA00000409F.l.20.1	74394	1	0	0	0	0	0	0	0
2472	RTA00000418F.l.06.1	73317	1	0	0	0	0	0	0	0
2473	RTA00000346F.o.22.1	7381	2	6	0	0	0	0	0	0
2474	RTA00000129A.k.22.1	79639	1	0	0	0	0	0	0	0
2476	RTA00000418F.m.22.1	74567	1	0	0	0	0	0	0	0
2477	RTA00000413F.c.12.1	65334	1	0	0	0	0	0	0	0
2479	RTA00000418F.g.20.1	74626	1	0	0	0	0	0	0	0
2480	RTA00000413F.d.15.1	64943	1	0	0	0	0	0	0	0
2483	RTA00000412F.c.10.1	76372	1	0	0	0	0	0	0	0
2484	RTA00000122A.j.17.1	62736	1	0	0	0	0	0	0	0
2489	RTA00000418F.j.19.1	78399	1	0	0	0	0	0	0	0
2490	RTA00000137A.p.12.1	80614	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2492	RTA00000418F.p.10.1	75323	1	0	0	0	0	0	0	0
2493	RTA00000408F.k.12.1	77246	1	0	0	0	0	0	0	0
2494	RTA00000137A.j.11.4	79752	1	0	0	0	0	0	0	0
2496	RTA00000419F.n.24.1	65995	1	0	0	0	0	0	0	0
2497	RTA00000418F.l.03.1	79058	1	0	0	0	0	0	0	0
2499	RTA00000419F.m.13.1	79052	1	0	0	0	0	0	0	0
2500	RTA00000418F.j.14.1	32623	1	1	0	0	0	0	0	0
2501	RTA00000403F.a.10.1	73952	1	0	0	0	0	0	0	0
2502	RTA00000420F.a.21.1	66241	1	0	0	0	0	0	0	0
2503	RTA00000127A.e.6.1	5885	4	2	0	0	0	0	0	0
2504	RTA00000405F.g.21.2	38966	2	0	0	0	0	0	0	0
2505	RTA00000405F.g.21.1	38966	2	0	0	0	0	0	0	0
2506	RTA00000419F.m.06.1	75749	1	0	0	0	0	0	0	0
2507	RTA00000423F.g.03.1	38007	2	0	0	0	0	0	0	0
2509	RTA00000418F.f.03.1	78911	1	0	0	0	0	0	0	0
2512	RTA00000120A.c.20.1	43235	1	1	0	0	0	1	0	0
2513	RTA00000138A.m.15.1	41603	1	1	0	0	0	0	0	0
2514	RTA00000408F.f.14.2	73024	1	0	0	0	0	0	0	0
2515	RTA00000418F.p.20.1	78023	1	0	0	0	0	0	0	0
2516	RTA00000423F.e.21.1	66961	1	0	0	0	0	0	0	0
2517	RTA00000419F.j.22.1	73525	1	0	0	0	0	0	0	0
2518	RTA00000410F.d.18.1	75458	1	0	0	0	0	0	0	0
2519	RTA00000403F.b.24.1	78838	1	0	0	0	0	0	0	0
2521	RTA00000410F.e.09.1	76093	1	0	0	0	0	0	0	0
2524	RTA00000353R.h.10.1	39498	2	0	0	0	0	0	0	0
2526	RTA00000411F.d.21.1	74794	1	0	0	0	0	0	0	0
2527	RTA00000340F.m.04.1	19406	2	1	0	0	0	0	0	0
2528	RTA00000411F.n.09.1	78962	1	0	0	0	0	0	0	0
2529	RTA00000127A.h.22.2	13155	2	3	0	0	0	0	0	0
2530	RTA00000420F.e.09.1	66325	1	0	0	0	0	0	0	0
2531	RTA00000405F.p.03.1	11346	3	3	0	0	0	0	0	0
2532	RTA00000419F.a.18.1	78484	1	0	0	0	0	0	0	0
2535	RTA00000121A.n.23.1	26981	2	0	0	0	0	0	0	0
2536	RTA00000121A.n.15.1	40849	1	1	0	0	0	0	0	0
2537	RTA00000403F.i.23.1	11364	4	2	0	0	0	0	0	0
2538	RTA00000405F.a.03.1	39065	2	0	0	0	0	0	0	0
2540	RTA00000419F.p.08.1	65560	1	0	0	0	0	0	0	0
2541	RTA00000126A.n.6.2	79917	1	0	0	0	0	0	0	0
2542	RTA00000413F.c.03.1	64527	1	0	0	1	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2543	RTA00000422F.k.24.1	39118	2	0	0	0	0	0	0	0
2544	RTA00000412F.c.17.1	75620	1	0	0	0	0	0	0	0
2546	RTA00000347F.g.08.1	23121	3	0	0	0	0	0	0	0
2547	RTA00000419F.o.06.1	64643	1	0	0	0	0	0	0	0
2548	RTA00000340R.j.07.1	38954	2	0	0	0	0	0	0	0
2549	RTA00000423F.j.02.1	38617	2	0	0	0	0	0	0	0
2550	RTA00000419F.c.04.1	63749	1	0	0	0	0	0	0	0
2551	RTA00000411F.a.01.1	74524	1	0	0	0	0	0	0	0
2552	RTA00000406F.f.05.1	22961	2	1	0	0	0	0	1	0
2553	RTA00000410F.n.05.1	77830	1	0	0	0	0	0	0	0
2554	RTA00000404F.e.06.1	39315	2	0	0	0	0	0	0	0
2556	RTA00000411F.c.03.1	79280	1	0	0	0	0	0	0	0
2562	RTA00000405F.l.07.1	38636	2	0	0	0	0	0	0	0
2564	RTA00000411F.n.06.1	73886	1	0	0	0	0	0	0	0
2565	RTA00000422F.k.15.1	19253	2	0	0	0	0	0	0	0
2566	RTA00000406F.h.16.1	38618	2	0	0	0	0	0	0	0
2567	RTA00000419F.f.24.1	18717	1	1	0	0	0	0	0	0
2568	RTA00000411F.d.18.1	76063	1	0	0	0	0	0	0	0
2571	RTA00000408F.d.15.1	78467	1	0	0	0	0	0	0	0
2572	RTA00000339F.b.22.1	6867	7	3	0	0	0	0	0	0
2574	RTA00000411F.n.02.1	78049	1	0	0	0	0	0	0	0
2575	RTA00000419F.b.17.1	63261	1	0	0	0	0	0	0	0
2577	RTA00000130A.e.20.1	79502	1	0	0	0	0	0	0	0
2579	RTA00000411F.i.13.1	66138	1	0	0	0	0	0	0	0
2580	RTA00000420F.e.20.1	64762	1	0	0	0	0	0	0	0
2581	RTA00000126A.p.23.2	80915	1	0	0	0	0	0	0	0
2583	RTA00000406F.g.08.1	37963	2	0	0	0	0	0	0	0
2584	RTA00000409F.a.08.1	74978	1	0	0	0	0	0	0	0
2585	RTA00000406F.d.24.1	37997	2	0	0	0	0	0	0	0
2588	RTA00000418F.i.12.1	78971	1	0	0	0	0	0	0	0
2589	RTA00000121A.h.19.1	80334	1	0	0	0	0	0	0	0
2590	RTA00000419F.b.10.1	78566	1	0	0	0	0	0	0	0
2591	RTA00000406F.m.10.1	38004	2	0	0	0	0	0	0	0
2592	RTA00000406F.o.05.1	37894	2	0	0	0	0	0	0	0
2593	RTA00000408F.b.04.2	39933	2	0	0	0	0	0	0	0
2594	RTA00000411F.k.04.1	65407	1	0	0	0	0	0	0	0
2596	RTA00000134A.l.9.1	81814	1	0	0	0	0	0	0	0
2598	RTA00000418F.k.04.1	75864	1	0	0	0	0	0	0	0
2601	RTA00000419F.p.18.1	63002	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2603	RTA00000419F.a.24.1	79290	1	0	0	0	0	0	0	0
2605	RTA00000129A.e.14.1	80053	1	0	0	0	0	0	0	0
2606	RTA00000404F.a.01.1	19251	2	0	0	0	0	0	0	0
2609	RTA00000408F.n.16.2	73720	1	0	0	0	0	0	0	0
2613	RTA00000412F.l.14.1	62792	1	0	0	0	0	0	0	0
2614	RTA00000129A.b.6.2	39111	2	0	0	0	0	0	0	0
2615	RTA00000406F.n.12.1	37517	2	0	0	0	0	0	0	0
2616	RTA00000418F.e.03.1	73442	1	0	0	0	0	0	0	0
2618	RTA00000403F.g.03.1	23537	2	1	0	0	0	0	0	0
2619	RTA00000412F.p.06.1	65485	1	0	0	0	0	0	0	0
2620	RTA00000419F.b.21.1	65366	1	0	0	0	0	0	0	0
2623	RTA00000351R.j.16.1	64773	1	0	0	0	0	0	0	0
2625	RTA00000419F.f.18.1	64047	1	0	0	0	0	0	0	0
2626	RTA00000423F.i.16.1	38604	2	0	0	0	0	0	0	0
2628	RTA00000411F.f.04.1	64526	1	0	0	0	0	0	0	0
2629	RTA00000125A.c.17.1	80619	1	0	0	0	0	0	0	0
2630	RTA00000404F.g.08.1	38980	2	0	0	0	0	0	0	0
2631	RTA00000423F.c.13.1	39059	2	0	0	0	0	0	0	0
2634	RTA00000404F.k.15.1	18225	2	0	0	0	0	0	0	0
2636	RTA00000339F.l.12.1	7711	4	1	0	0	0	0	0	0
2637	RTA00000406F.b.01.1	39006	2	0	0	0	0	0	0	0
2638	RTA00000407F.c.08.1	37549	2	0	0	0	0	0	0	0
2640	RTA00000403F.b.05.1	74300	1	0	0	0	0	0	0	0
2644	RTA00000408F.j.05.2	73878	1	0	0	0	0	0	0	0
2646	RTA00000419F.c.14.1	65727	1	0	0	0	0	0	0	0
2650	RTA00000346F.h.24.1	4379	9	2	0	0	0	0	0	0
2651	RTA00000420F.b.02.1	64013	1	0	0	0	0	0	0	0
2652	RTA00000413F.b.24.1	65117	1	0	0	0	0	0	0	0
2653	RTA00000412F.d.08.1	75328	1	0	0	0	0	0	0	0
2655	RTA00000419F.m.18.1	76014	1	0	0	0	0	0	0	0
2656	RTA00000419F.l.24.1	74628	1	0	0	0	0	0	0	0
2657	RTA00000408F.c.06.1	78619	1	0	0	0	0	0	0	0
2658	RTA00000405F.h.21.2	39072	2	0	0	0	0	0	0	0
2660	RTA00000405F.g.05.2	38987	2	0	0	0	0	0	0	0
2661	RTA00000411F.f.20.1	63501	1	0	0	0	0	0	0	0
2663	RTA00000420F.d.19.1	43146	1	1	0	0	0	0	0	0
2664	RTA00000195R.a.06.1	35265	2	0	1	0	0	0	0	0
2665	RTA00000123A.f.2.1	80379	1	0	0	0	0	0	0	0
2666	RTA00000411F.j.11.1	66154	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2671	RTA00000419F.j.03.1	77578	1	0	0	0	0	0	0	0
2673	RTA00000423F.h.11.1	38977	2	0	0	0	0	0	0	0
2674	RTA00000413F.b.17.1	21704	1	2	0	0	0	0	0	0
2677	RTA00000423F.f.03.1	63852	1	0	0	0	0	0	0	0
2678	RTA00000419F.e.10.1	63225	1	0	0	0	0	0	0	0
2680	RTA00000403F.d.02.1	39224	2	0	0	0	0	0	0	0
2682	RTA00000418F.j.20.1	77101	1	0	0	0	0	0	0	0
2690	RTA00000356R.h.05.1	35052	2	0	1	0	0	0	0	0
2692	RTA00000340F.i.15.1	26815	1	0	0	0	0	0	0	0
2694	RTA00000345F.c.12.1	23824	2	1	0	0	0	0	0	0
2696	RTA00000412F.o.03.1	65039	1	0	0	0	0	0	0	0
2697	RTA00000409F.d.16.1	76090	1	0	0	0	0	0	0	0
2700	RTA00000408F.j.17.2	78935	1	0	0	0	0	0	0	0
2701	RTA00000126A.j.15.2	40425	2	0	0	0	0	0	0	0
2705	RTA00000410F.b.17.1	77458	1	0	0	0	0	0	0	0
2706	RTA00000419F.l.22.1	78444	1	0	0	0	0	0	0	0
2708	RTA00000422F.f.22.1	38703	2	0	0	0	0	0	0	0
2711	RTA00000418F.c.05.1	76475	1	0	0	0	0	0	0	0
2712	RTA00000418F.p.21.1	78068	1	0	0	0	0	0	0	0
2714	RTA00000340F.i.08.1	12005	2	1	0	0	0	0	0	0
2715	RTA00000410F.o.04.1	79018	1	0	0	0	0	0	0	0
2716	RTA00000411F.l.16.1	16122	1	3	0	0	0	0	0	0
2717	RTA00000411F.j.03.1	66263	1	0	0	0	0	0	0	0
2718	RTA00000126A.k.24.1	39428	2	0	0	0	0	0	0	0
2720	RTA00000120A.m.10.3	81376	1	0	0	0	0	0	0	0
2721	RTA00000419F.f.16.1	64679	1	0	0	0	0	0	0	0
2722	RTA00000408F.c.23.1	42261	1	1	0	0	0	0	0	0
2725	RTA00000136A.h.6.1	81620	1	0	0	0	0	0	0	0
2730	RTA00000418F.e.20.1	73741	1	0	0	0	0	0	0	0
2732	RTA00000405F.l.03.1	38580	2	0	0	0	0	0	0	0
2733	RTA00000418F.m.02.1	74550	1	0	0	0	0	0	0	0
2735	RTA00000406F.c.05.1	22077	3	0	1	0	0	0	0	0
2737	RTA00000411F.k.21.1	65349	1	0	0	0	0	0	0	0
2741	RTA00000418F.i.06.1	75151	1	0	0	0	0	0	0	0
2742	RTA00000423F.a.03.1	26796	2	0	0	0	0	0	0	0
2744	RTA00000423F.k.21.2	37499	2	0	0	0	0	0	0	0
2746	RTA00000404F.c.18.1	38982	2	0	0	0	0	0	0	0
2749	RTA00000411F.g.24.1	65233	1	0	0	0	0	0	0	0
2751	RTA00000405F.m.07.1	37733	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2752	RTA00000411F.j.07.1	66963	1	0	0	0	0	0	0	0
2754	RTA00000353R.h.04.1	17123	4	0	0	0	0	0	0	0
2755	RTA00000408F.f.10.2	75309	1	0	0	0	0	0	0	0
2757	RTA00000405F.o.03.1	37575	2	0	0	0	0	0	0	0
2758	RTA00000413F.b.18.1	39873	2	0	0	0	0	0	0	0
2764	RTA00000408F.c.08.1	73473	1	0	0	0	0	0	0	0
2766	RTA00000410F.c.06.1	77784	1	0	0	0	1	0	0	0
2768	RTA00000405F.b.08.1	39182	2	0	0	0	0	0	0	0
2769	RTA00000409F.l.24.1	73174	1	0	0	0	0	0	0	0
2770	RTA00000406F.j.06.1	38952	2	0	0	0	0	0	0	0
2771	RTA00000423F.h.03.1	37903	2	0	0	0	0	0	0	0
2773	RTA00000121A.k.22.1	79523	1	0	0	0	0	0	0	0
2775	RTA00000411F.m.06.1	24195	2	1	0	0	0	0	0	0
2776	RTA00000126A.b.9.1	81279	1	0	0	0	0	0	0	0
2779	RTA00000404F.l.05.1	38671	2	0	0	0	0	0	0	0
2785	RTA00000419F.p.10.1	41448	1	1	0	0	0	0	0	0
2786	RTA00000120A.c.19.1	81016	1	0	0	0	0	0	0	0
2792	RTA00000411F.k.14.1	63987	1	0	0	0	0	0	0	0
2793	RTA00000420F.e.05.1	63908	1	0	0	0	0	0	0	0
2796	RTA00000128A.j.10.1	80085	1	0	0	0	0	0	0	0
2797	RTA00000412F.f.10.2	65405	1	0	0	0	0	0	0	0
2799	RTA00000422F.k.17.1	38955	2	0	0	0	0	0	0	0
2801	RTA00000347F.h.10.1	22779	3	0	0	0	0	0	0	0
2803	RTA00000419F.l.02.1	75736	1	0	0	0	0	0	0	0
2805	RTA00000418F.b.20.1	73560	1	0	0	0	0	0	0	0
2808	RTA00000408F.n.05.2	77883	1	0	0	0	0	0	0	0
2809	RTA00000419F.o.09.1	66396	1	0	0	0	0	0	0	0
2814	RTA00000422F.o.08.2	26832	2	0	0	0	0	0	0	0
2817	RTA00000418F.m.18.1	76479	1	0	0	0	0	0	0	0
2818	RTA00000347F.e.20.1	39911	2	0	0	0	0	0	0	0
2819	RTA00000419F.e.23.1	65772	1	0	0	0	0	0	0	0
2826	RTA00000411F.g.05.1	64664	1	0	0	0	0	0	0	0
2827	RTA00000404F.h.10.1	37148	2	0	0	0	0	0	0	0
2828	RTA00000422F.n.14.1	26787	2	0	0	0	0	0	0	0
2830	RTA00000120A.m.13.3	80608	1	0	0	0	0	0	0	0
2831	RTA00000412F.i.03.1	65617	1	0	0	0	0	0	0	0
2832	RTA00000418F.l.02.1	39316	2	0	0	0	0	0	0	0
2834	RTA00000411F.j.04.1	66219	1	0	0	0	0	0	0	0
2839	RTA00000404F.a.18.1	36267	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2840	RTA00000408F.l.14.1	12001	2	3	0	0	0	0	0	0
2841	RTA00000405F.d.10.1	39000	2	0	0	0	0	0	0	0
2843	RTA00000418F.h.23.1	75153	1	0	0	0	0	0	0	0
2845	RTA00000418F.j.11.1	73853	1	0	0	0	0	0	0	0
2846	RTA00000408F.o.13.1	74895	1	0	0	0	0	0	0	0
2847	RTA00000419F.o.07.1	14059	1	0	0	0	0	0	0	0
2848	RTA00000419F.n.17.1	63186	1	0	0	0	0	0	0	0
2849	RTA00000403F.f.15.1	22768	3	0	0	0	0	0	0	0
2850	RTA00000408F.d.03.1	22768	3	0	0	0	0	0	0	0
2852	RTA00000346F.f.02.1	62757	1	0	0	0	0	0	0	0
2854	RTA00000413F.i.21.1	64066	1	0	0	0	0	0	0	0
2856	RTA00000419F.h.21.1	64828	1	0	0	0	0	0	0	0
2865	RTA00000121A.a.2.1	81843	1	0	0	0	0	0	0	0
2866	RTA00000527F.g.13.1	36035	2	0	0	0	0	0	0	0
2869	RTA00000426F.h.11.1	75479	1	0	0	0	0	0	0	0
2874	RTA00000522F.b.22.1	75181	1	0	0	0	0	0	0	0
2877	RTA00000522F.a.23.1	38613	2	0	0	0	0	0	0	0
2879	RTA00000523F.b.02.1	65163	1	0	0	0	0	0	0	0
2880	RTA00000425F.j.14.1	73397	1	0	0	0	0	0	0	0
2883	RTA00000522F.e.16.1	75283	1	0	0	0	0	0	0	0
2886	RTA00000523F.h.17.1	65586	1	0	0	0	0	0	0	0
2888	RTA00000522F.p.07.1	76888	1	0	0	0	0	0	0	0
2889	RTA00000522F.n.08.1	76343	1	0	0	0	0	0	0	0
2890	RTA00000425F.c.06.1	78041	1	0	0	0	0	0	0	0
2891	RTA00000427F.b.23.1	64297	1	0	0	0	0	0	0	0
2892	RTA00000527F.p.02.1	36844	2	0	0	0	0	0	0	0
2893	RTA00000427F.d.08.1	63967	1	0	0	0	0	0	0	0
2895	RTA00000426F.m.07.1	63504	1	0	0	0	0	0	0	0
2896	RTA00000427F.c.10.1	65478	1	0	0	0	0	0	0	0
2899	RTA00000424F.m.15.1	73759	1	0	0	0	0	0	0	0
2900	RTA00000426F.f.11.1	63102	1	0	0	0	0	0	0	0
2902	RTA00000426F.f.20.1	65134	1	0	0	0	0	0	0	0
2907	RTA00000527F.i.19.2	38089	2	0	0	0	0	0	0	0
2912	RTA00000523F.e.18.1	62898	1	0	0	0	0	0	0	0
2913	RTA00000527F.k.21.1	36051	2	0	0	0	0	0	0	0
2916	RTA00000522F.n.02.1	74959	1	0	0	0	0	0	0	0
2919	RTA00000425F.f.19.1	32635	1	1	0	0	0	0	0	0
2920	RTA00000528F.e.23.1	19242	3	0	0	0	0	0	0	0
2921	RTA00000522F.n.16.1	26769	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2922	RTA00000427F.c.20.1	26527	1	0	0	0	0	0	0	0
2923	RTA00000527F.k.06.1	12469	3	1	0	0	0	0	0	0
2925	RTA00000523F.i.06.1	66341	1	0	0	0	0	0	0	0
2926	RTA00000427F.f.21.1	36853	2	0	0	0	0	0	0	0
2927	RTA00000427F.j.19.1	41395	1	1	0	0	0	0	0	0
2928	RTA00000522F.b.01.1	75691	1	0	0	0	0	0	0	0
2929	RTA00000424F.i.24.1	79101	1	0	0	0	0	0	0	0
2930	RTA00000523F.c.01.1	65710	1	0	0	0	0	0	0	0
2931	RTA00000427F.b.15.1	66891	1	0	0	0	0	0	0	0
2934	RTA00000522F.j.15.2	76535	1	0	0	0	0	0	0	0
2937	RTA00000426F.f.19.1	66701	1	0	1	0	0	0	0	0
2940	RTA00000523F.i.22.1	64688	1	0	0	0	0	0	0	0
2942	RTA00000425F.i.17.1	43213	1	1	0	0	0	0	0	0
2945	RTA00000425F.p.12.1	73219	1	0	0	0	0	0	0	0
2946	RTA00000427F.j.07.1	64819	1	0	0	0	0	0	0	0
2948	RTA00000527F.i.05.2	37481	2	0	0	0	0	0	0	0
2951	RTA00000523F.k.01.1	41437	1	1	0	0	0	0	0	0
2952	RTA00000425F.j.11.1	76667	1	0	0	0	0	0	0	0
2953	RTA00000424F.b.22.4	72971	1	0	0	0	0	0	0	0
2955	RTA00000525F.a.03.1	36786	2	0	0	0	0	0	0	0
2956	RTA00000527F.i.21.2	37490	2	0	0	0	0	0	0	0
2957	RTA00000424F.a.24.4	73951	1	0	0	0	0	0	0	0
2958	RTA00000522F.k.14.1	74280	1	0	0	0	0	0	0	0
2959	RTA00000522F.n.05.1	73260	1	0	0	0	0	0	0	0
2960	RTA00000523F.c.18.1	66179	1	0	0	0	0	0	0	0
2961	RTA00000523F.b.13.1	66330	1	0	0	0	0	0	0	0
2963	RTA00000527F.p.16.1	23798	2	1	0	0	0	0	0	0
2964	RTA00000425F.c.20.1	73581	1	0	0	0	0	0	0	0
2965	RTA00000424F.i.21.1	73482	1	0	0	0	0	0	0	0
2966	RTA00000523F.j.19.1	65910	1	0	0	0	0	0	0	0
2968	RTA00000424F.b.22.1	72971	1	0	0	0	0	0	0	0
2969	RTA00000527F.b.18.1	37469	2	0	0	0	0	0	0	0
2973	RTA00000525F.e.16.1	36837	2	0	0	0	0	0	0	0
2975	RTA00000522F.d.08.1	74284	1	0	0	0	0	0	0	0
2978	RTA00000527F.g.07.1	37488	2	0	0	0	0	0	0	0
2980	RTA00000525F.b.05.1	21116	2	1	0	0	0	0	0	0
2981	RTA00000425F.n.05.1	73965	1	0	0	0	0	0	0	0
2982	RTA00000523F.d.18.1	64072	1	0	0	0	0	0	0	0
2983	RTA00000525F.a.02.1	37454	2	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
2985	RTA00000426F.h.09.1	78797	1	0	0	0	0	0	0	0
2988	RTA00000427F.g.05.1	63138	1	0	0	0	0	0	0	0
2989	RTA00000424F.m.12.1	77675	1	0	0	0	0	0	0	0
2995	RTA00000427F.h.12.1	36894	2	0	0	0	0	0	0	0
2996	RTA00000523F.c.15.1	36935	2	0	0	0	0	0	0	0
2997	RTA00000427F.k.17.1	64965	1	0	0	0	0	0	0	0
2999	RTA00000424F.c.14.3	76614	1	0	0	0	0	0	0	0
3000	RTA00000522F.k.10.2	77619	1	0	0	0	0	0	0	0
3001	RTA00000424F.m.22.1	72943	1	0	0	0	0	0	0	0
3002	RTA00000527F.h.17.1	37799	2	0	0	0	0	0	0	0
3003	RTA00000527F.c.22.1	37496	2	0	0	0	0	0	0	0
3004	RTA00000425F.k.22.1	78123	1	0	0	0	0	0	0	0
3005	RTA00000424F.m.14.1	77491	1	0	0	0	0	0	0	0
3006	RTA00000522F.k.19.1	32625	1	1	0	0	0	0	0	0
3007	RTA00000523F.i.18.1	64463	1	0	0	0	0	0	0	0
3008	RTA00000425F.j.22.1	73882	1	0	0	0	0	0	0	0
3009	RTA00000527F.g.23.1	37538	2	0	0	0	0	0	0	0
3010	RTA00000426F.m.24.1	63943	1	0	0	0	0	0	0	0
3012	RTA00000425F.d.21.1	78920	1	0	0	0	0	0	0	0
3014	RTA00000424F.d.04.3	76505	1	0	0	0	0	0	0	0
3015	RTA00000424F.d.04.1	76505	1	0	0	0	0	0	0	0
3016	RTA00000427F.c.12.1	66995	1	0	0	0	0	0	0	0
3018	RTA00000527F.l.13.1	36904	2	0	0	0	0	0	0	0
3019	RTA00000522F.h.13.1	40823	1	1	0	0	0	0	0	0
3020	RTA00000424F.l.19.1	75454	1	0	0	0	0	0	0	0
3023	RTA00000427F.a.06.1	66550	1	0	0	0	0	0	0	0
3024	RTA00000525F.c.19.1	38159	2	0	0	0	0	0	0	0
3025	RTA00000523F.f.06.1	62871	1	0	0	0	0	0	0	0
3026	RTA00000424F.h.10.1	72925	1	0	0	0	0	0	0	0
3027	RTA00000522F.a.12.1	33515	1	1	0	0	0	0	0	0
3028	RTA00000522F.h.01.1	75010	1	0	0	0	0	0	0	0
3030	RTA00000425F.e.21.1	77203	1	0	0	0	0	0	0	0
3031	RTA00000523F.f.07.1	62799	1	0	0	0	0	0	0	0
3033	RTA00000424F.j.12.1	73827	1	0	0	0	0	0	0	0
3035	RTA00000523F.d.12.1	64888	1	0	0	0	0	0	0	0
3036	RTA00000523F.e.10.1	62878	1	0	0	0	0	0	0	0
3037	RTA00000425F.f.11.1	79275	1	0	0	0	0	0	0	0
3038	RTA00000426F.m.18.1	62974	1	0	0	0	0	0	0	0
3041	RTA00000522F.g.15.1	76536	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
3042	RTA00000522F.n.12.1	74117	1	0	0	0	0	0	0	0
3044	RTA00000424F.d.10.3	73110	1	0	0	0	0	0	0	0
3048	RTA00000527F.c.04.1	23090	3	0	0	0	0	0	0	0
3050	RTA00000527F.h.21.1	37630	2	0	0	0	0	0	0	0
3051	RTA00000425F.c.07.1	76042	1	0	0	0	0	0	0	0
3053	RTA00000525F.c.15.1	7692	2	0	0	0	0	0	0	0
3054	RTA00000424F.d.22.3	76189	1	0	0	0	0	0	0	0
3055	RTA00000523F.h.12.1	65745	1	0	0	0	0	0	0	0
3056	RTA00000522F.g.22.1	77504	1	0	0	0	0	0	0	0
3059	RTA00000522F.j.12.2	74341	1	0	0	0	0	0	0	0
3060	RTA00000523F.i.08.1	65099	1	0	0	0	0	0	0	0
3062	RTA00000425F.j.20.1	26760	1	0	0	0	0	0	0	0
3064	RTA00000427F.f.24.1	64572	1	0	0	0	0	0	0	0
3065	RTA00000527F.a.13.1	37740	2	0	0	0	0	0	0	0
3069	RTA00000424F.a.09.4	77833	1	0	0	0	0	0	0	0
3071	RTA00000525F.f.07.1	37500	2	0	0	0	0	0	0	0
3072	RTA00000424F.j.07.1	79211	1	0	0	0	0	0	0	0
3073	RTA00000424F.m.10.1	34251	1	1	0	0	0	0	0	0
3075	RTA00000522F.g.06.1	78221	1	0	0	0	0	0	0	0
3076	RTA00000424F.h.03.1	74447	1	0	0	0	0	0	0	0
3077	RTA00000424F.n.06.1	74737	1	0	0	0	0	0	0	0
3078	RTA00000427F.c.22.1	63990	1	0	0	0	0	0	0	0
3079	RTA00000424F.k.12.1	77666	1	0	0	0	0	0	0	0
3080	RTA00000425F.f.02.1	76982	1	0	0	0	0	0	0	0
3081	RTA00000427F.h.11.1	26494	1	0	0	0	0	0	0	0
3082	RTA00000425F.j.16.1	75631	1	0	0	0	0	0	0	0
3084	RTA00000427F.f.17.1	63803	1	0	0	0	0	0	0	0
3085	RTA00000522F.o.18.1	76366	1	0	0	0	0	0	0	0
3086	RTA00000427F.j.22.1	66367	1	0	0	0	0	0	0	0
3087	RTA00000426F.p.10.1	65845	1	0	0	0	0	0	0	0
3088	RTA00000522F.m.02.1	76834	1	0	0	0	0	0	0	0
3091	RTA00000425F.e.15.1	75921	1	0	0	0	0	0	0	0
3094	RTA00000424F.n.13.1	74942	1	0	0	0	0	0	0	0
3095	RTA00000424F.g.14.1	74879	1	0	0	0	0	0	0	0
3096	RTA00000426F.e.17.1	64089	1	0	0	0	0	0	0	0
3100	RTA00000427F.g.19.1	64611	1	0	0	0	0	0	0	0
3102	RTA00000522F.c.01.1	74938	1	0	0	0	0	0	0	0
3103	RTA00000522F.g.17.1	76486	1	0	0	0	0	0	0	0
3104	RTA00000523F.j.17.1	63610	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
3105	RTA00000522F.n.14.1	73410	1	0	0	0	0	0	1	0
3107	RTA00000523F.e.20.1	65164	1	0	0	0	0	0	0	0
3108	RTA00000424F.c.15.3	73533	1	0	0	0	0	0	0	0
3109	RTA00000426F.p.09.1	66665	1	0	0	0	0	0	0	0
3110	RTA00000522F.p.09.1	75204	1	0	0	0	0	0	0	0
3111	RTA00000426F.m.21.1	64915	1	0	0	0	0	0	0	0
3112	RTA00000425F.j.21.1	77373	1	0	0	0	0	0	0	0
3114	RTA00000523F.h.21.1	41440	1	1	0	0	0	0	0	0
3115	RTA00000427F.h.24.1	65193	1	0	0	0	0	0	0	0
3116	RTA00000425F.f.24.1	40841	1	1	0	0	0	0	0	0
3117	RTA00000425F.m.03.1	76045	1	0	0	0	0	0	0	0
3118	RTA00000426F.m.08.1	63781	1	0	0	0	0	0	0	0
3119	RTA00000523F.d.24.1	64799	1	0	0	0	0	0	0	0
3120	RTA00000523F.c.14.1	66015	1	0	0	0	0	0	0	0
3121	RTA00000523F.b.20.1	66492	1	0	0	0	0	0	0	0
3122	RTA00000522F.h.07.1	75149	1	0	0	0	0	0	0	0
3123	RTA00000527F.g.10.1	37820	2	0	0	0	0	0	0	0
3126	RTA00000427F.i.22.1	63199	1	0	0	0	0	0	0	0
3128	RTA00000527F.n.07.1	15939	2	2	0	0	0	0	0	0
3129	RTA00000425F.e.09.1	75550	1	0	0	0	0	0	0	0
3130	RTA00000427F.h.02.1	63652	1	0	0	0	0	0	0	0
3131	RTA00000426F.f.16.1	65613	1	0	0	0	0	0	0	0
3132	RTA00000425F.i.21.1	75305	1	0	0	0	0	0	0	0
3133	RTA00000427F.k.19.1	62851	1	0	0	0	0	0	0	0
3135	RTA00000426F.g.16.1	41446	1	1	0	0	0	0	0	0
3136	RTA00000527F.l.05.1	13016	4	0	0	1	1	0	0	0
3137	RTA00000426F.m.02.1	66237	1	0	0	0	0	0	0	0
3140	RTA00000522F.l.22.1	75801	1	0	0	0	0	0	0	0
3141	RTA00000427F.h.19.1	63047	1	0	0	0	0	0	0	0
3143	RTA00000522F.g.21.1	77310	1	0	0	0	0	0	0	0
3145	RTA00000522F.g.20.1	77688	1	0	0	0	0	0	0	0
3148	RTA00000425F.k.20.1	74048	1	0	0	0	0	0	0	0
3150	RTA00000522F.b.07.1	78634	1	0	0	0	0	0	0	0
3151	RTA00000426F.g.19.1	63672	1	0	0	0	0	0	0	0
3152	RTA00000525F.d.19.1	36860	2	0	0	0	0	0	0	0
3154	RTA00000427F.d.10.1	40685	1	1	0	0	0	0	0	0
3157	RTA00000424F.a.05.4	77976	1	0	0	0	0	0	0	0
3159	RTA00000424F.a.05.1	77976	1	0	0	0	0	0	0	0
3160	RTA00000522F.l.15.1	74691	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
3161	RTA00000425F.e.02.1	76143	1	0	0	0	0	0	0	0
3162	RTA00000525F.c.11.1	37895	2	0	0	0	0	0	0	0
3164	RTA00000522F.c.14.1	75449	1	0	0	0	0	0	0	0
3165	RTA00000424F.m.08.1	19402	1	2	0	0	0	0	0	0
3166	RTA00000527F.f.18.1	37577	2	0	0	0	0	0	0	0
3168	RTA00000522F.a.06.1	73662	1	0	0	0	0	0	0	0
3171	RTA00000522F.d.23.1	73868	1	0	0	0	0	0	0	0
3174	RTA00000523F.j.10.1	63384	1	0	0	0	0	0	0	0
3175	RTA00000527F.p.08.1	36013	2	0	0	0	0	0	0	0
3177	RTA00000426F.f.17.1	66334	1	0	0	0	0	0	0	0
3178	RTA00000523F.j.21.1	36925	2	0	0	0	0	0	0	0
3183	RTA00000523F.a.01.1	74923	1	0	0	0	0	0	0	0
3185	RTA00000427F.j.06.1	63676	1	0	0	0	0	0	0	0
3186	RTA00000424F.m.04.1	79017	1	0	0	0	0	0	0	0
3187	RTA00000523F.i.17.1	65779	1	0	0	0	0	0	0	0
3190	RTA00000525F.c.18.1	24208	2	1	0	0	0	0	0	0
3191	RTA00000527F.e.09.1	37521	2	0	0	0	0	0	0	0
3192	RTA00000424F.j.08.1	73972	1	0	0	0	0	0	0	0
3194	RTA00000527F.c.09.1	64859	1	0	0	0	0	0	0	0
3197	RTA00000523F.c.03.1	36913	2	0	0	0	0	0	0	0
3198	RTA00000427F.k.21.1	62880	1	0	0	0	0	0	0	0
3200	RTA00000427F.d.09.1	66486	1	0	0	0	0	0	0	0
3201	RTA00000426F.n.17.1	66572	1	0	0	0	0	0	0	0
3204	RTA00000426F.m.03.1	66480	1	0	0	0	0	0	0	0
3205	RTA00000424F.h.06.1	77552	1	0	0	0	0	0	0	0
3206	RTA00000425F.d.06.1	77660	1	0	0	0	0	0	0	0
3207	RTA00000427F.e.12.1	62813	1	0	0	0	0	0	0	0
3210	RTA00000426F.n.23.1	18176	1	0	0	0	0	0	0	0
3211	RTA00000522F.m.19.1	41544	1	1	0	0	0	0	0	0
3212	RTA00000522F.a.05.1	32611	1	1	0	0	0	0	0	0
3213	RTA00000427F.i.09.1	65916	1	0	0	0	0	0	0	0
3214	RTA00000424F.j.09.1	74387	1	0	0	0	0	0	0	0
3215	RTA00000424F.n.11.1	73874	1	0	0	0	0	0	0	0
3217	RTA00000527F.e.13.1	37588	2	0	0	0	0	0	0	0
3219	RTA00000425F.j.19.1	77925	1	0	0	0	0	0	0	0
3220	RTA00000522F.g.12.1	78783	1	0	0	0	0	0	0	0
3221	RTA00000523F.a.07.1	75804	1	0	0	0	0	0	0	0
3222	RTA00000425F.e.19.1	73409	1	0	0	0	0	0	0	0
3223	RTA00000425F.n.19.1	78324	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
3228	RTA00000427F.k.07.1	63742	1	0	0	0	0	0	0	0
3231	RTA00000522F.a.17.1	79032	1	0	0	0	0	0	0	0
3232	RTA00000527F.l.19.1	36856	2	0	0	0	0	0	0	0
3233	RTA00000424F.i.11.1	41569	1	1	0	0	0	0	0	0
3235	RTA00000424F.d.19.3	73180	1	0	0	0	0	0	0	0
3236	RTA00000522F.j.09.2	78522	1	0	0	0	0	0	0	0
3237	RTA00000424F.m.24.1	77045	1	0	0	0	0	0	0	0
3238	RTA00000522F.j.19.2	76224	1	0	0	0	0	0	0	0
3242	RTA00000527F.j.12.2	37503	2	0	0	0	0	0	0	0
3243	RTA00000522F.g.11.1	75432	1	0	0	0	0	0	0	0
3244	RTA00000522F.k.02.2	77622	1	0	0	0	0	0	0	0
3245	RTA00000427F.e.13.1	66080	1	0	0	0	0	0	0	0
3246	RTA00000426F.f.18.1	63271	1	0	0	0	0	0	0	0
3247	RTA00000427F.a.12.1	63377	1	0	0	0	0	0	0	0
3248	RTA00000424F.b.23.4	77322	1	0	0	0	0	0	0	0
3252	RTA00000427F.f.02.1	36822	2	0	0	0	0	0	0	0
3254	RTA00000424F.i.15.1	78043	1	0	0	0	0	0	0	0
3256	RTA00000522F.m.03.1	79194	1	0	0	0	0	0	0	0
3257	RTA00000522F.a.20.1	74070	1	0	0	0	0	0	0	0
3258	RTA00000424F.b.15.4	74958	1	0	0	0	0	0	0	0
3259	RTA00000527F.g.14.1	37532	2	0	0	0	0	0	0	0
3260	RTA00000522F.d.06.1	74809	1	0	0	0	0	0	0	0
3262	RTA00000427F.e.10.1	64599	1	0	0	0	0	0	0	0
3263	RTA00000527F.c.16.1	22908	3	0	0	0	0	0	0	0
3265	RTA00000523F.f.17.1	63984	1	0	0	0	0	0	0	0
3267	RTA00000527F.p.24.1	36832	2	0	0	0	0	0	0	0
3268	RTA00000425F.n.17.1	78304	1	0	0	0	0	0	0	0
3270	RTA00000425F.e.07.1	75992	1	0	0	0	0	0	0	0
3272	RTA00000523F.h.08.1	62893	1	0	0	0	0	0	0	0
3273	RTA00000522F.o.10.1	78798	1	0	0	0	0	0	0	0
3274	RTA00000425F.l.10.1	26893	1	0	0	0	0	0	0	0
3275	RTA00000427F.f.16.1	64122	1	0	0	0	0	0	0	0
3278	RTA00000425F.i.10.1	78736	1	0	0	0	0	0	0	0
3279	RTA00000426F.m.12.1	63740	1	0	0	0	0	0	0	0
3280	RTA00000527F.g.12.1	37746	2	0	0	0	0	0	0	0
3283	RTA00000425F.i.18.1	42255	1	1	0	0	0	0	0	0
3285	RTA00000424F.j.13.1	74485	1	0	0	0	0	0	0	0
3289	RTA00000424F.k.10.1	73232	1	0	0	0	0	0	0	0
3290	RTA00000522F.i.07.2	78377	1	0	0	0	0	0	0	0

SEQ ID NO:	Sequence Name	cluster	lib 1 clones	lib 2 clones	lib 15 clones	lib 16 clones	lib 17 clones	lib 18 clones	lib 19 clones	lib 20 clones
3292	RTA00000522F.b.08.1	26915	1	0	0	0	0	0	0	0
3293	RTA00000522F.l.08.1	78781	1	0	0	0	0	0	0	0
3294	RTA00000525F.a.14.1	37566	2	0	0	0	0	0	0	0
3295	RTA00000424F.g.08.1	74928	1	0	0	0	0	0	0	0
3296	RTA00000425F.l.09.1	75251	1	0	0	0	0	0	0	0
3297	RTA00000522F.o.20.1	74853	1	0	0	0	0	0	0	0
3298	RTA00000527F.j.04.2	11809	3	1	0	0	0	0	0	0
3300	RTA00000523F.c.13.1	40668	1	1	0	0	0	0	0	0
3301	RTA00000427F.i.21.1	65540	1	0	0	0	0	0	0	0
3303	RTA00000522F.h.02.1	74947	1	0	0	0	0	0	0	0
3304	RTA00000522F.g.10.1	74294	1	0	0	0	0	0	0	0
3308	RTA00000425F.k.16.1	75282	1	0	0	0	0	0	0	0
3309	RTA00000525F.b.09.1	23472	2	1	0	0	0	0	0	0
3310	RTA00000522F.j.08.2	76613	1	0	0	0	0	0	0	0
3312	RTA00000523F.f.19.1	34169	1	1	0	0	0	0	0	0
3313	RTA00000425F.j.18.1	75561	1	0	0	0	0	1	0	0
3314	RTA00000426F.m.04.1	36865	2	0	0	0	0	0	0	0
3315	RTA00000527F.g.21.1	36028	2	0	0	0	0	0	0	0
3317	RTA00000525F.a.22.1	36848	2	0	0	0	0	0	0	0
3318	RTA00000522F.p.22.1	73322	1	0	0	0	0	0	0	0
3319	RTA00000424F.d.12.2	74342	1	0	0	0	0	0	0	0
3320	RTA00000424F.g.24.1	79156	1	0	0	0	0	0	0	0
3321	RTA00000427F.a.10.1	65370	1	0	0	0	0	0	0	0
3322	RTA00000426F.h.20.1	23187	3	0	0	0	0	0	0	0
3323	RTA00000424F.d.12.3	74342	1	0	0	0	0	0	0	0
3324	RTA00000425F.c.03.1	74643	1	0	0	0	0	0	0	0
3325	RTA00000523F.f.16.1	26522	1	0	0	0	0	0	0	0
3326	RTA00000427F.f.15.1	66734	1	0	0	0	0	0	0	0
3329	RTA00000522F.p.18.1	76376	1	0	0	0	0	0	0	0
3337	RTA00000522F.g.18.1	73226	1	0	0	0	0	0	0	0
3339	RTA00000522F.h.05.1	73358	1	0	0	0	0	0	0	0
3341	RTA00000425F.n.16.1	18265	1	0	0	0	0	0	0	0
3342	RTA00000527F.l.21.1	36439	2	0	0	0	0	0	0	0
3345	RTA00000424F.d.17.3	73958	1	0	0	0	0	0	0	0
3346	RTA00000523F.j.02.1	62853	1	0	0	0	0	0	0	0

No clones corresponding to the colon-specific polynucleotides in the table above were present in any of Libraries 3, 4, 8, 9, 12, 13, 14, or 15. The polynucleotide provided

above can be used as markers of cells of colon origin, and find particular use in reference arrays, as described above.

Example 26: Identification of Contiguous Sequences Having a Polynucleotide of the Invention

The novel polynucleotides were used to screen publicly available and proprietary databases to determine if any of the polynucleotides of SEQ ID NOS: 845-3346 would facilitate identification of a contiguous sequence, *e.g.*, the polynucleotides would provide sequence that would result in 5' extension of another DNA sequence, resulting in production of a longer contiguous sequence composed of the provided polynucleotide and the other DNA sequence(s). Contiging was performed using the Gelmerge application (default settings) of GCG from the Univ. of Wisconsin.

Using these parameters, 146 contiged sequences were generated. These contiged sequences are provided as SEQ ID NOS:5951-6096(see Table 17). The contiged sequences can be correlated with the sequences of SEQ ID NOS:845-3346 upon which the contiged sequences are based by, for example, identifying those sequences of SEQ ID NOS: 845-3346 and the contiged sequences of SEQ ID NOS: 5951-6096that share the same clone name in Table 17.

The contiged sequences (SEQ ID NO: 5951-6096) thus represent longer sequences that encompass a polynucleotide sequence of the invention. The contiged sequences were then translated in all three reading frames to determine the best alignment with individual sequences using the BLAST programs as described above for SEQ ID NOS: 845-3346 and the validation sequences "SEQ ID NOS:3347-5950." Again the sequences were masked using the XBLAST program for masking low complexity as described above in Example 1 (Table 18). Several of the contiged sequences were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 36). Thus the invention encompasses fragments, fusions, and variants of such polynucleotides that retain biological activity associated with the protein family and/or functional domain identified herein.

Table 36 Profile hits using contiged sequences

SEQ ID NO	Biological Activity (Profile)	Start	Stop	Score	Direction	Sequence Name
5955	7tm_2	71	915	8090	for	RTA00000399F.o.01.1
5964	7tm_2	101	919	8475	rev	RTA00000341F.m.21.1
6018	7tm_2	3	963	9431	for	RTA00000192AF.h.19.1
6041	7tm_2	214	1073	8528	rev	RTA00000192AF.f.3.1
6052	ANK	546	629	4920	for	RTA00000190AF.f.5.1
5964	asp	126	1067	6620	rev	RTA00000341F.m.21.1.
6085	asp	112	1094	6553	for	RTA00000418F.i.06.1
6087	asp	347	1028	5981	for	RTA00000339F.b.02.1
6041	ATPases	113	781	5690	for	RTA00000192AF.f.3.1
6083	ATPases	1	348	15955	for	RTA00000401F.m.07.1
6085	ATPases	110	823	6782	for	RTA00000418F.i.06.1
6087	ATPases	338	874	5832	for	RTA00000339F.b.02.1
5969	protkinase	59	685	5791	for	RTA00000182AF.c.5.1
6061	protkinase	75	1035	5405	for	RTA00000181AF.p.12.3
6081	protkinase	25	546	5107	rev	RTA00000118A.n.5.1
6092	protkinase	14	422	5103	rev	RTA00000419F.k.05.1
6096	protkinase	89	755	5499	for	RTA00000404F.m.17.2
5964	Wnt_dev_sign	3	948	11036	for	RTA00000341F.m.21.1

All stop/start sequences are provided in the forward direction.

Descriptions of the profiles for the indicated protein families and functional domains are provided in Example 3 above.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Deposit Information:

The following materials were deposited with the American Type Culture Collection: CMCC = (Chiron Master Culture Collection)

Cell Lines Deposited with ATCC

Cell Line	Deposit Date	ATCC Accession No.	CMCC Accession No.
KM12L4-A	March 19, 1998	CRL-12496	11606
Km12C	May 15, 1998	CRL-12533	11611
MDA-MB-231	May 15, 1998	CRL-12532	10583
MCF-7	October 9, 1998	CRL-12584	10377

cDNA Libraries Deposited with ATCC

cDNA Library No. Deposit Date ATCC Accession No.	cDNA Library ES21 January 22, 1999 ATCC No.	cDNA Library ES22 January 22, 1999 ATCC No.	cDNA Library ES23 January 22, 1999 ATCC No.
Clone Names	M00001575D:G05	M00001364A:E11	M00001489B:A06
	M00001460A:A03	M00001694C:H10	M00001585A:D06
	M00001655C:E04	M00003841D:E03	M00001637B:E07
	M00001676C:C11	M00004176D:B12	M00001529D:H02
	M00001679D:D05	M00001387B:E02	M00001500C:C08
	M00001546B:C05	M00004282B:A04	M00001483B:D03
	M00001453B:E10	M00001376B:F03	M00001623C:H07
		M00001445D:A06	M00003975B:F03
		M00001399C:H12	
		M00004208D:H08	

cDNA Library No. Deposit Date ATCC Accession No.	cDNA Library ES24 January 22, 1999 ATCC No.	cDNA Library ES25 January 22, 1999 ATCC No.	cDNA Library ES26 January 22, 1999 ATCC No.
Clone Names	M00003987D:D06	M00001675D:B08	M00001479C:F10
	M00004073A:H12	M00001589B:E12	M00003842D:F08
	M00004104B:F11	M00001607D:A11	M00003901A:C09
	M00004237D:D08	M00001636A:E07	M00003982A:B06

M00004111D:B07	M00001530A:B12	M00003824A:A06
M00004138B:B11	M00001495B:B08	M00003845D:C03
M00001391C:C04	M00001487C:F01	M00003856A:B07
M00001448D:E12	M00001644B:D06	M00004104B:A02
M00001450A:B03	M00003751C:A04	M00004110C:E03
M00001451B:F01		

In addition, libraries of selected clones were deposited. The details of these deposits are provided in Tables 37-40.

This deposit is provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby.

Table 37. Clones Deposited on January 22, 1999

cDNA LIbrary Ref. ATCC No	Library ES17 207064	Library ES18 207065	Library ES19 207066
Clone Names	M00001601A:E09	M00001594A:D06	M00003906A:F04
	M00001368A:D07	M00001613D:H10	M00003908A:F12
	M00003917A:D02	M00001596D:E10	M00003914A:G09
	M00001673A:A04	M00001592C:G04	M00003915C:H04
	M00003868B:G11	M00001599D:A09	M00003905D:B08
	M00003917C:D03	M00001619B:A09	M00003908C:G09
	M00003791C:E09	M00001593B:E11	M00003914B:A11
	M00003870A:C05	M00001605A:E06	M00003916C:C05
	M00003922A:D02	M00001608A:D03	M00003959A:A03
	M00003861C:H02	M00001616C:A02	M00003905D:C08
	M00003931B:A11	M00001617A:D06	M00003908D:D12
	M00001679D:B05	M00001595C:E01	M00003901B:H04
	M00001679C:D05	M00001616C:A11	M00004031A:E01
	M00001687A:G01	M00001608C:E11	M00004029C:C12
	M00003945A:E09	M00001610C:E06	M00003911A:F10
	M00003908A:H09	M00001612B:D11	M00003914C:F09
	M00001649B:G12	M00001618B:E05	M00003963D:B05
	M00003813D:H12	M00001621C:C10	M00003986C:E09
	M00004087C:D03	M00001647A:H08	M00004031A:F07
	M00004269B:C08	M00001631D:B10	M00003907C:C02
	M00004348A:A02	M00001608D:E09	M00003911B:F08
	M00001679C:D01	M00001641B:C10	M00003914C:H05
	M00001490A:E11	M00001641D:E02	M00003918C:C12
	M00001387A:E10	M00001630D:H10	M00003914C:C02

M00001397B:G03	M00001585C:D10	M00003914A:E04
M00001441D:E04	M00001560A:H10	M00003903B:D03
M00001352C:G09	M00001573B:C06	M00003905A:F09
M00001370D:A12	M00001660C:D11	M00003867C:E11
M00001387B:A06	M00001641C:C05	M00003870B:B08
M00001397C:A10	M00001578B:B05	M00003879D:A08
M00001536D:G02	M00001587C:C10	M00003891D:B10
M00003895C:A10	M00001590B:C07	M00003901C:A08
M00001464B:B03	M00001554A:E04	M00003903C:C04
M00004370A:G05	M00001570C:G06	M00003905A:F10
M00001490B:H11	M00001576A:B09	M00003906C:D06
M00001530B:D10	M00001582A:H01	M00003907D:A12
M00001579C:E09	M00001582B:E12	M00003905C:G11
M00001587A:H03	M00001615B:F07	M00003914D:D10
M00001457C:H12	M00001571C:A04	M00003972A:G09
M00001535C:E01	M00001573D:D10	M00003975D:C06
M00001561D:C05	M00001576A:F11	M00003905C:B02
M00001589A:C01	M00001579C:G05	M00003907D:F11
M00001664D:G07	M00001582D:A02	M00003914A:G06
M00001565A:H09	M00001589B:E07	M00003914D:E03
M00001381C:B08	M00001575B:B02	M00003972C:F08
M00001395C:F11	M00001578C:G06	M00003976C:D06
M00001429D:F11	M00001591A:B08	M00003907C:C04
M00001449A:F01	M00001607A:F11	M00003905B:C06
M00001391C:H02	M00001579C:E06	M00004088C:A12
M00001429D:H12	M00001661C:F11	M00004103C:D04
M00001450A:G11	M00001650B:C10	M00004107A:D01
M00001344B:F12	M00001654C:E04	M00004110A:E04
M00001391D:C06	M00001656B:A08	M00004062A:H06
M00003971A:A06	M00001662C:B02	M00004075D:C10
M00001346A:E04	M00001656B:D05	M00004081D:H09
M00001455C:G07	M00001661C:F10	M00004089A:B08
M00001402D:F02	M00001663A:C11	M00004103D:F10
M00001438D:C06	M00001669A:C10	M00004107B:B04
M00001349B:G05	M00001651B:B12	M00004032C:B02
M00001389C:A08	M00001653B:E06	M00004078C:F04
M00001439B:A10	M00001659C:F02	M00004038B:H10
M00001455B:A09	M00001661B:F03	M00004089A:E02
M00001441B:D11	M00001663C:F10	M00004096B:F05
M00001453A:B01	M00001669A:G12	M00004104C:H12
M00001456D:E08	M00001674D:C10	M00004110D:A10
M00001399A:C03	M00001651B:E06	M00004036D:F02
M00004496C:H03	M00001651C:C05	M00004088C:E04
M00004135D:G02	M00001657C:C07	M00004104D:A04
M00004692A:E07	M00001662A:C12	M00004107D:E12
M00004374D:E10	M00001663D:C06	M00004115D:D08
M00004405D:C04	M00001590B:C05	M00003846A:D03
M00004312B:H07	M00001483C:G06	M00004072C:F08
M00003976C:A10	M00001653A:G07	M00004039B:G08
M00004043A:D02	M00001625B:C10	M00003986D:D02

M00004081C:H06	M00001626C:D12	M00003914A:B07
M00004050D:A06	M00001634D:D02	M00003914D:B02
M00001361B:C07	M00001641C:C06	M00003971B:B07
M00004341B:G03	M00001642D:F02	M00003978C:A03
M00001342B:E01	M00001647B:E04	M00003983B:C08
M00004064D:A11	M00001632B:E05	M00004033D:D07
M00004087A:G08	M00001639A:C11	M00004072D:H12
M00004344B:H04	M00001642D:G10	M00004077B:H11
M00004497A:H03	M00001624A:G11	M00004080A:F01
M00001338C:E10	M00001626C:G08	M00004092C:B03
M00001366D:E12	M00001672D:D04	M00004037B:C04
M00001390D:E03	M00001639A:H06	M00004073C:D04
M00001413B:H09	M00001662C:A04	M00004081A:A08
M00004271B:B06	M00001641B:B01	M00004085B:B05
M00004151D:E03	M00001673C:A02	M00004090C:C07
M00001660B:C04	M00001650A:A12	M00004086D:B09
M00003802D:B11	M00001659D:D03	M00004088D:B03
M00001579C:E08	M00001661B:B05	M00004090C:C10
M00001557D:C08	M00001671D:E10	M00004102C:D09
M00003779B:E12	M00001652D:A06	M00004105C:E09
M00001638A:D10	M00001654C:D05	M00004035A:G10
M00003794A:B03	M00001656A:B07	M00003906A:H07
M00001616C:F07	M00001647B:C09	M00004083B:G03
M00001679A:F01	M00001635A:C06	M00001675B:E02
M00001604C:E09	M00001482D:A04	M00003793C:D09
M00001653B:E09	M00001485C:B10	M00003762B:H09
M00001585A:F07	M00001457D:A07	M00001694C:F12
M00003811D:A12	M00001461A:E05	M00001678D:C11
M00001653C:F12	M00001477A:G07	M00001677D:B07
M00001679D:F06	M00001479D:H03	M00001677B:A02
M00003751D:B02	M00001482C:D02	M00001675B:H03
M00003801A:B10	M00001484D:G05	M00003808D:D04
M00003844C:A08	M00001459B:D03	M00003752B:C02
M00001636C:C01	M00001464B:C11	M00003819D:B11
M00001669C:B01	M00001511A:A05	M00001677D:B02
M00003755A:A09	M00001477B:C02	M00001694C:G04
M00003798D:H08	M00001471A:D04	M00003789C:F06
M00001444C:D05	M00001485C:H10	M00001678C:C06
M00004040B:F10	M00001485D:E05	M00001675B:D02
M00001355A:C12	M00001487C:G03	M00003750C:H05
M00001401A:H07	M00001514A:B04	M00001694A:B12
M00001393B:B09	M00001530C:G10	M00001677B:H06
M00001409D:F11	M00001534A:G06	M00001675C:G01
M00001387B:H07	M00001539A:C12	M00001675B:C01
M00001394C:C11	M00001547A:F11	M00003857B:F07
M00001344A:H07	M00001550D:A04	M00003812B:D07
M00001490C:D07	M00001460A:F07	M00001694B:B08
M00001352C:F06	M00001472C:A01	M00001677B:E06
M00001476D:G03	M00001481B:A07	M00004037A:E04
M00001399C:D09	M00001456D:F05	M00003870A:H01

M00001347C:G08	M00001456D:G11	M00003842C:D11
M00001453D:G12	M00001477D:F10	M00003828B:F09
M00001382A:F04	M00001481A:G06	M00003856C:H09
M00001392D:H04	M00001464A:B03	M00003851A:C10
M00001429C:G12	M00001469A:G11	M00003841C:E04
M00001454A:C11	M00001478B:D07	M00003837C:G08
M00001517B:G08	M00001473A:C11	M00003828B:E07
M00001535A:D02	M00001457A:G03	M00003772C:B12
M00001352A:E12	M00001669B:G02	M00001677D:F03
M00001381B:F06	M00001479D:G06	M00001678B:B12
M00004117A:D11	M00001473D:B11	M00001678D:G03
M00004217C:D03	M00001475A:A12	M00001675C:F01
M00004270A:F11	M00001460A:G07	M00003809A:H04
M00003996A:A06	M00001464A:D03	M00003771D:G05
M00004056B:D09	M00001473D:G01	M00001678A:F05
M00004142A:B12	M00001476D:C05	M00001677B:B06
M00001396D:B03	M00001484A:A10	M00003794A:E12
M00001370D:E12	M00001457C:F02	M00003771B:E05
M00001390C:C11	M00001459B:A12	M00001678A:A11
M00003989A:H11	M00001464A:E07	M00003805B:C04
M00001426A:A09	M00001467A:B03	M00001680B:E10
M00004498D:D05	M00001514A:B08	M00001679B:H07
M00001391B:G12	M00001464A:B07	M00003904D:B12
M00001391D:D10	M00001579A:C03	M00003856C:B08
M00001376B:A02	M00001517A:G08	M00003858D:G06
M00001405B:D07	M00001530B:G09	M00003870B:F04
M00001368A:A03	M00001538A:F12	M00003871C:B05
M00001392D:B11	M00001540C:B03	M00003875A:C04
M00003900D:B10	M00001547A:F06	M00003901B:A09
M00001494B:C01	M00001550A:F07	M00003901C:D03
M00001352C:A05	M00001567B:G11	M00003904C:B06
M00001408B:G06	M00001572A:A10	M00003901C:F09
M00004252C:E03	M00001575B:G01	M00003904D:B10
M00003901C:A03	M00001487D:C11	M00003850D:H11
M00004071D:A10	M00001577B:A03	M00003902B:D06
M00001377B:H01	M00001539D:E10	M00003879A:C01
M00003939A:A02	M00001587A:F05	M00003877D:G05
M00004250D:D10	M00001560A:F03	M00003881D:C12
M00004290A:B03	M00001569B:G11	M00003903A:H09
M00003911D:B04	M00001573A:A06	M00003905A:A06
M00004128B:G01	M00001575D:A10	M00003875D:D09
M00004142A:D08	M00001583A:D01	M00003879B:A06
M00003977A:E04	M00001587A:F08	M00003823D:G05
M00004236C:D10	M00001590B:B02	M00003763A:C01
M00004388B:A08	M00001553A:E07	M00003903B:C02
M00004409B:A11	M00001560A:H06	M00003905A:E07
M00003965A:B11	M00001589C:A11	M00003867A:D12
M00003988A:E10	M00001538A:C08	M00003857C:C09
M00004138A:H09	M00001531A:H03	M00003829C:D10
M00003933C:D06	M00001548A:G01	M00003839D:E02

M00004193C:G11	M00001531A:H07	M00003841C:F03
M00004039C:C01	M00001542A:E04	M00003903D:C06
M00003924B:D04	M00001487A:F10	M00003852D:E08
M00004375C:D01	M00001503C:G05	M00003845D:A09
	M00001511A:G08	M00003824A:G10
	M00001539A:H12	M00003841C:F06
	M00001542A:F06	M00003848A:C09
	M00001549A:F01	M00003857C:F11
	M00001514A:A12	M00003816C:C01
	M00001516A:D05	M00003843A:E08
	M00001546C:C07	M00003850A:F06
	M00001549A:H11	M00003813B:A11
	M00001538A:D03	M00003855C:F10
	M00001544A:C09	M00003850D:B05
	M00001546B:F12	M00003841D:F06
	M00001550A:D09	M00003858B:G05
	M00001487B:F02	M00003854D:A12
	M00001513A:G07	M00003857C:G01
	M00001530A:F12	M00003816C:E09
	M00001538A:D12	M00003813A:G04
	M00001587A:G06	M00003850D:A05
	M00001551A:D04	
	M00001485B:C03	

Table 38. Clones Deposited on January 22, 1999

cDNA Ref No.; Clone Names in Library	cDNA Library Ref ES20 ATCC No. 207067	cDNA Ref No. ES27 ATCC No. 207074	cDNA Library Ref ATCC No. 207075
	M00004891D:A07	**M00001623B:G07**	**M00001550D:H02**
	M00004118B:C11	**M00001619D:G05**	**M00001549C:D02**
	M00004105A:B10	**M00001616C:C09**	**M00001549A:A09**
	M00004099A:F11	**M00001615C:F03**	**M00001548A:B11**
	M00004037C:D07	**M00001614D:D09**	**M00001546C:G10**
	M00004033D:C05	**M00001608B:A03**	**M00001544C:C06**
	M00003983D:A09	**M00001607D:F07**	**M00003820B:C05**
	M00004029B:H08	**M00001623D:C10**	**M00001543A:H12**
	M00004927A:A02	**M00001599B:E09**	**M00001540C:B10**
	M00003983C:F10	**M00001632C:C09**	**M00001552B:G05**
	M00003980B:C06	**M00001605C:D12**	**M00001543C:F01**
	M00004033D:B07	**M00001625D:C07**	**M00001552D:G08**
	M00004034C:E08	**M00001629B:E06**	**M00001554B:B07**
	M00005100B:H07	**M00001594A:B12**	**M00001555A:B01**
	M00005136A:D10	**M00001632C:A02**	**M00001557A:F01**
	M00005173D:H02	**M00001567C:H12**	**M00001558A:E11**
	M00004891D:C11	**M00001635C:A03**	**M00001561C:E11**
	M00004101A:F07	**M00001636C:H09**	**M00001571D:B11**
	M00003982B:B06	**M00001638A:E07**	**M00001563B:D11**
	M00004108C:E01	**M00001639A:F10**	**M00001569C:B06**
	M00005136D:B07	**M00001656C:G08**	**M00001539B:H06**

cDNA Ref No.;	cDNA Library Ref ES20	cDNA Ref No. ES27	cDNA Library Ref
	M00004118D:A11	M00001632A:F12	M00001571B:E03
	M00005102C:C01	M00001557A:D02	M00001561D:C11
	M00005177C:A01	M00001529B:C04	M00001487C:D06
	M00004927C:H11	M00001534B:C12	M00001454B:D08
	M00005174D:B02	M00001535D:C01	M00003772D:E10
	M00004027A:D06	M00001536D:A12	M00001573C:D03
	M00005217A:G10	M00001540B:C09	M00001454D:E05
	M00003984A:B06	M00001540D:D02	M00001455D:F09
	M00003851C:D07	M00001541C:B07	M00001457C:C11
	M00003959C:G06	M00001546B:B02	M00001459B:C09
	M00005100B:G11	M00001575B:C09	M00001460A:E01
	M00005213C:G01	M00001554B:C07	M00001460C:H02
	M00003982B:H07	M00001578D:C04	M00001456A:H02
	M00004029C:B03	M00001557C:H07	M00001477B:F04
	M00004033D:G06	M00001558B:D08	M00003845D:B04
	M00004091B:H09	M00001560D:A03	M00001488A:E01
	M00003959D:A04	M00001561C:F06	M00001492D:A11
	M00004030D:B06	M00001564D:C09	M00001496C:G10
	M00004034C:C06	M00003748B:F02	M00001499A:A05
	M00004030C:D12	M00001570D:A03	M00001500A:B02
	M00003982C:H10	M00001660C:B12	M00001500D:E10
	M00003971C:F09	M00001577B:H02	M00001513D:A03
	M00004031B:A06	M00001548A:A08	M00001528A:C11
	M00003966B:D02	M00003868B:D12	M00001528C:H04
	M00004028B:G08	M00001718D:F07	M00001531B:E09
	M00004031C:H10	M00003829C:A11	M00001463A:F06
	M00004076D:B09	M00003832B:E01	M00003755A:B03
	M00004092D:B11	M00003842B:D09	M00001653B:G07
	M00003981C:F05	M00003845A:H12	M00001654D:G11
	M00004031D:F05	M00003847B:G03	M00001656B:A07
	M00004097B:D03	M00003847C:E09	M00001664B:D06
	M00003986D:G07	M00003853D:G08	M00001664C:H10
	M00004033B:C02	M00003828A:E04	M00001680B:C01
	M00004037B:A04	M00003867C:H09	M00001681A:F03
	M00004092C:B12	M00003822A:F02	M00001684B:G03
	M00005140D:G09	M00003868C:H10	M00001771A:A07
	M00004897D:G05	M00003871A:A05	M00003774C:D02
	M00004960B:D12	M00003879C:G10	M00003754D:D02
	M00005134C:G04	M00003880C:F10	M00001640B:F03
	M00005139A:F01	M00003881D:D06	M00003763B:H01
	M00005176A:C12	M00003884D:G07	M00003812C:A05
	M00005178A:A07	M00003887A:A06	M00003803C:D09
	M00005212A:A02	M00003889A:D10	M00003801B:B10
	M00005229D:H07	M00003889D:B09	M00003798D:E03
	M00004115C:H04	M00003858D:F12	M00003773B:G01
	M00004687A:C03	M00003774B:B08	M00003771A:G10
	M00004900C:E11	M00001680D:D02	M00001452A:E07
	M00004695B:E04	M00001528A:F09	M00004029B:F11

cDNA Ref No.;	cDNA Library Ref ES20	cDNA Ref No. ES27	cDNA Library Ref
	M00005134D:A06	M00003748A:B07	M00003751B:A05
	M00004103B:B07	M00001655A:F06	M00001609B:A11
	M00005177A:B06	M00003750A:D01	M00001573D:F10
	M00005178A:A08	M00003761D:E02	M00001579C:B11
	M00004104D:B05	M00003763D:E10	M00001579C:H10
	M00004117B:G01	M00003768A:E02	M00001579D:G07
	M00004900D:B10	M00003829B:G03	M00001583B:E10
	M00005134D:H03	M00003772A:D07	M00001586D:E02
	M00005173C:A02	M00001661B:C08	M00001587D:A10
	M00005177A:H09	M00003778A:D08	M00001589A:D12
	M00005178B:H01	M00003799A:D09	M00001590C:H08
	M00005216C:B09	M00003800A:C09	M00001651B:A11
	M00003826B:E11	M00003804A:H04	M00001597A:E12
	M00001596A:G06	M00003806D:G05	M00001649C:B10
	M00005100B:D02	M00003808C:B05	M00001614A:E06
	M00005137A:E01	M00003811A:E03	M00001615C:D02
	M00004119A:A06	M00003815D:H09	M00001621D:D03
	M00004891D:E07	M00003818B:G12	M00001623D:G03
	M00004958B:D01	M00003769B:D03	M00001624A:F09
	M00005102C:F09	M00001390A:A09	M00001624C:A06
	M00005136D:C01	M00001432A:E06	M00001630B:A11
	M00005174D:H02	M00001381A:D02	M00001634B:C10
	M00005177C:B04	M00001383A:G04	M00001639D:B07
	M00005218B:D09	M00001384C:E03	M00001573D:F04
	M00004102C:F03	M00001384C:F12	M00001595B:A09
	M00004114B:D09	M00001384D:H07	M00004156B:A12
	M00004119D:A07	M00001385B:F10	M00004319D:G09
	M00004895C:G05	M00001385C:H11	M00004096A:G02
	M00004235A:A12	M00001386A:C02	M00004101C:G08
	M00005134B:E01	M00001372C:F07	M00004102A:H02
	M00004115C:G03	M00001389D:G11	M00004108A:A09
	M00005175B:H04	M00001371D:G01	M00004111D:D11
	M00005214B:D11	M00001392C:D10	M00004115D:C08
	M00004102D:B05	M00001392D:H06	M00004118D:E08
	M00004115A:B12	M00001397B:B09	M00004121C:F06
	M00004119D:H06	M00001398A:G03	M00004131B:H09
	M00004897D:F03	M00001400A:F06	M00004141D:A09
	M00004960B:A09	M00001410B:G05	M00004090A:F09
	M00005134C:E11	M00001413A:F02	M00004146A:C08
	M00005138B:D12	M00001415B:E09	M00004078B:A11
	M00005176A:A05	M00001425A:C11	M00004176B:E08
	M00005214C:A09	M00001386A:D11	M00004188C:A09
	M00004102C:D01	M00001354C:B06	M00004233C:H09
	M00004960B:A08	M00001339D:G02	M00004241D:F11
	M00001476D:A09	M00001660A:C12	M00004246C:A09
	M00001572A:B06	M00001528A:A01	M00004247C:C12
	M00005217D:F12	M00001343D:C04	M00004248B:E08
	M00005233A:G08	M00001347B:E01	M00004257C:H06
	M00005236B:F10	M00001348A:D04	M00004260D:C12

cDNA Ref No.;	cDNA Library Ref ES20	cDNA Ref No. ES27	cDNA Library Ref
	M00005259B:C01	M00001349C:C05	M00004295B:D02
	M00005254D:B08	M00001350A:D06	M00004040D:F01
	M00005259C:B05	M00001352D:C05	M00004142D:E10
	M00001575A:D06	M00001380C:E05	M00003853D:D03
	M00005259D:H08	M00001354B:B10	M00003860D:H07
	M00003813C:D08	M00001380C:F02	M00003878C:E04
	M00001530D:E06	M00001354C:C10	M00003879A:G05
	M00004891B:B12	M00001355B:G11	M00003880B:C08
	M00001596B:C11	M00001356D:F06	M00003881A:D09
	M00004300C:H09	M00001360D:E11	M00003881C:G09
	M00001486D:D12	M00001361C:H11	M00003901B:A05
	M00001585D:F03	M00001362C:A10	M00003904D:D10
	M00001596B:D09	M00001363C:H02	M00003905C:G10
	M00001570D:E06	M00001366D:G02	M00003906B:F12
	M00001582C:E01	M00001369A:H12	M00003909A:H04
	M00001586C:E06	M00001352D:D02	M00004091B:D11
	M00001593B:D10	M00001485D:B10	M00003963A:E03
	M00001595C:H11	M00001457B:E03	M00004353C:H07
	M00001596B:H05	M00001457C:C12	M00003919A:A10
	M00001576A:C11	M00001458C:E01	M00003938A:B04
	M00001596C:F09	M00001462B:A10	M00003939C:F04
	M00001567A:H05	M00001464D:F06	M00003946D:C11
	M00001585D:D11	M00001467D:H05	M00003979A:F03
	M00004688A:A02	M00001468B:H06	M00003985C:F01
	M00004927A:E06	M00001505C:H01	M00003997B:G07
	M00005229D:H09	M00001470A:H01	M00003860D:A01
	M00004117B:A12	M00001457A:B07	M00004035A:A04
	M00004187D:G09	M00001479B:A01	M00004042D:H02
	M00005173B:F01	M00001469D:D02	M00004073B:B01
	M00005218A:G05	M00001487A:A05	M00003946A:H10
	M00004118A:H08	M00001352C:H02	M00001423D:A09
	M00005134A:D11	M00001488D:C10	M00004314B:G07
	M00005176C:C09	M00001490C:C12	M00001405D:D11
	M00005230D:F06	M00001493B:D09	M00001408A:H04
	M00005234D:B04	M00001504D:D11	M00001408D:D04
	M00005101C:E09	M00001376B:C06	M00001411D:F05
	M00004206A:E02	M00001506B:D09	M00001412A:E04
	M00001570C:A05	M00001511B:C06	M00001413A:F03
	M00005231A:H04	M00001476B:F10	M00001417B:C04
	M00005235A:A03	M00001450D:D04	M00001417D:A04
	M00004118B:B04	M00001433A:G07	M00001418B:F07
	M00005136D:D06	M00001470C:B10	M00001419D:C10
	M00005231C:B01	M00001437D:C04	M00001402B:F12
	M00004153B:B03	M00001447C:C01	M00001423A:G05
	M00004897C:D06	M00001448B:F06	M00001401C:H03
	M00005136D:G06	M00001449D:A06	M00001423D:D12
	M00005212B:A02	M00001433B:H11	M00001424B:H04
	M00005232A:C10	M00001451D:C10	M00001428B:A09
	M00004692A:H10	M00001452A:C07	M00001430A:A02

cDNA Ref No.;	cDNA Library Ref ES20	cDNA Ref No. ES27	cDNA Library Ref
	M00005101C:B09	M00001453C:A11	M00001432D:F05
	M00004144A:F04	M00001456B:C09	M00001438B:B09
	M00003852B:D11	M00001454B:G03	M00001445B:E04
	M00001660D:E05	M00001454B:G07	M00001445C:A08
	M00003808A:F09	M00001454C:C08	M00001446C:D09
	M00001656A:D10	M00001454C:F02	M00001448A:G09
	M00001671A:H06	M00001454D:D06	M00001449C:H12
	M00003809C:H07	M00001456B:F10	M00001422C:F12
	M00003853C:C06	M00001455D:A09	M00001352C:H10
	M00003860A:A08	M00001455D:A11	M00004375A:H01
	M00003822B:D08	M00001448D:F09	M00004380B:A05
	M00003845A:E12		M00004444B:D11
	M00003854C:C02		M00001338B:E02
	M00003860B:G09		M00001341A:F12
	M00003822B:G01		M00001344A:G07
	M00001670A:C11		M00001345A:G11
	M00003852A:B03		M00001345B:E10
	M00003829D:A11		M00001345C:B01
	M00003854C:F01		M00001346B:B07
	M00003856B:C04		M00001405B:E09
	M00003905A:H11		M00001352B:F04
	M00001530A:F11		M00001451C:E01
	M00003840B:E07		M00001361A:H07
	M00003905B:G03		M00001362B:H06
	M00003840B:E08		M00001372C:G12
	M00003855A:C12		M00001375B:G12
	M00003905B:H05		M00001376A:C05
	M00003826B:B04		M00001376B:A08
	M00003851C:B06		M00001377C:E12
	M00003853B:C08		M00001382B:F12
	M00003829A:F03		M00001385A:F12
	M00001638C:G01		M00001394A:E04
	M00003845D:B02		M00001395A:C09
	M00001653D:G07		M00001396A:H03
	M00001578B:A02		M00001350B:G11
	M00001590B:H10		
	M00001595C:A09		
	M00001596A:E07		
	M00001607A:B06		
	M00001607A:D10		
	M00001652C:B09		
	M00001671B:F02		
	M00001632C:D08		
	M00001638C:H07		
	M00001652D:B09		
	M00001614C:E11		
	M00001633B:B11		
	M00001651C:A04		
	M00001639D:G12		

cDNA Ref No.;	cDNA Library Ref ES20	cDNA Ref No. ES27	cDNA Library Ref
	M00001671C:F11		
	M00001638A:B04		
	M00001637C:H12		
	M00001669B:H06		
	M00001639D:F02		
	M00001590A:C08		
	M00001636A:C02		
	M00001614A:A04		
	M00001639D:G06		

Table 39. Library Deposited on January 22, 1999

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES29 ATCC No. 207076	cDNA Library Ref ES30 ATCC No. 207077
Clone Names in Library	M00001449D:B01	M00001594D:B08
	M00001476D:F03	M00001593A:B07
	M00001456C:B12	M00001594A:C01
	M00001469B:B01	M00001594A:D08
	M00001471A:B04	M00001594A:G09
	M00001472A:D08	M00001595C:B05
	M00001473A:A07	M00001594B:F12
	M00001473C:D09	M00001596D:E03
	M00001475B:C04	M00001594D:C03
	M00001475C:G11	M00001592C:F11
	M00001476A:D11	M00001590D:G07
	M00001476B:D10	M00001595D:A04
	M00001468A:C05	M00001595D:G03
	M00001476C:C11	M00001601A:A06
	M00001467A:H07	M00001590C:F10
	M00001477B:E02	M00001589B:B08
	M00001478B:H08	M00001589C:E06
	M00001479C:E01	M00001611B:A05
	M00001480A:D03	M00001601A:E02
	M00001480C:A05	M00001587A:D01
	M00001481A:H08	M00001591B:B12
	M00001481B:D09	M00001590B:G08
	M00001482A:H05	M00001592C:E05
	M00001482D:H11	M00001591B:B06
	M00001483C:G09	M00001591D:C07
	M00001485A:C05	M00001591D:F06
	M00001476B:F08	M00001592A:E02
	M00001460A:E11	M00001592A:H05
	M00001456C:C11	M00001592B:A04
	M00001457A:C05	M00001587A:B10
	M00001457A:G12	M00001609D:G10
	M00001458A:A11	M00005231D:B09
	M00001458C:D10	M00001614B:E08
	M00001458D:A01	M00005217C:C01
	M00001458D:A02	M00001587A:B01
	M00001458D:C11	M00001613D:B03

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES29 ATCC No. 207076	cDNA Library Ref ES30 ATCC No. 207077
	M00001458D:D01	M00001613A:F03
	M00001459B:C11	M00001611C:H11
	M00001468A:H10	M00001611C:C12
	M00001460A:C10	M00001611B:E06
	M00001485B:F05	M00001611B:A09
	M00001460A:H11	M00001610D:D05
	M00001461A:F05	M00001610B:C07
	M00001462A:D03	M00001610C:E07
	M00001464A:B02	M00001610A:E09
	M00001464A:E10	M00001601A:E12
	M00001465A:B12	M00001609B:C09
	M00001465A:C12	M00001608D:D11
	M00001465A:E10	M00001608B:A09
	M00001465A:G06	M00001607D:F06
	M00001466A:F08	M00001607B:C05
	M00001467A:C10	M00001606A:H09
	M00001460A:B12	M00001605A:H03
	M00001545A:B12	M00001605A:E09
	M00001535A:D10	M00001605A:A06
	M00001536A:F11	M00001604A:C11
	M00001537A:H05	M00001604A:C07
	M00001539A:E01	M00001604A:B08
	M00001539A:H02	M00001604A:A09
	M00001539B:G07	M00001610A:H05
	M00001539D:B10	M00005214B:A06
	M00001540D:E02	M00005228A:A09
	M00001541B:E05	M00001567A:B09
	M00001542A:G12	M00001561A:D01
	M00001485B:D09	M00001559A:C08
	M00001545A:B10	M00001559A:A11
	M00001533A:G05	M00001558A:G09
	M00001545A:F02	M00001555A:B12
	M00001545A:G05	M00001554A:A08
	M00001546A:D08	M00001552A:H10
	M00001548A:H04	M00001552A:F06
	M00001550A:E07	M00005231C:B07
	M00001551A:A11	M00005218D:G10
	M00001551A:D06	M00001570A:H01
	M00001551A:H06	M00005214D:D10
	M00001551D:H07	M00001570C:G03
	M00001552A:E10	M00005213C:A01
	M00001450A:B08	M00005212D:F08
	M00001544A:F05	M00005212A:D10
	M00001512A:G05	M00005211C:E09
	M00001483B:D04	M00005211A:E09
	M00001485B:H03	M00005210D:C09
	M00001485C:C08	M00005179D:B03
	M00001486B:D07	M00005179B:H02

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES29 ATCC No. 207076	cDNA Library Ref ES30 ATCC No. 207077
	M00001486B:E12	M00005177D:F09
	M00001487B:A11	M00005177C:G04
	M00001487B:E10	M00005177B:H02
	M00001507A:A11	M00001614D:B08
	M00001507A:B02	M00001615A:D06
	M00001507A:C05	M00005216B:D02
	M00001507A:E04	M00001579C:A01
	M00001534A:D03	M00001585B:C03
	M00001511A:G01	M00001585B:A06
	M00001533D:A08	M00001584D:H02
	M00001513A:F05	M00001584A:G03
	M00001514A:G03	M00001583D:B08
	M00001516A:D02	M00001583B:F02
	M00001516A:F06	M00001583A:F07
	M00001517A:B11	M00001583A:A05
	M00001529D:C05	M00001582D:F02
	M00001530A:A09	M00001582D:B01
	M00001530A:E10	M00001582A:A03
	M00001532A:C01	M00001579D:H09
	M00001532D:A06	M00001567D:B03
	M00001485B:D10	M00001579C:H06
	M00001511A:A02	M00001585B:F01
	M00004249D:B08	M00001579B:F04
	M00004185D:E04	M00001579A:E03
	M00004188D:G08	M00001578C:F05
	M00004197C:F03	M00001577D:H06
	M00004198B:D02	M00001577B:F10
	M00004204D:C03	M00001576C:G05
	M00004208B:F05	M00001575D:D12
	M00004208D:B10	M00001575D:B10
	M00004210B:B05	M00001575D:A02
	M00001362D:H01	M00001573B:G08
	M00004216D:D03	M00001573A:E01
	M00004167A:H03	M00001572A:B05
	M00004275A:B03	M00001571D:F05
	M00004285C:A08	M00001579D:F04
	M00004316A:G09	M00001636A:F08
	M00004465B:D04	M00001643B:E05
	M00004493B:D09	M00001642C:G02
	M00001347B:H04	M00001642A:F03
	M00001351C:B06	M00001641D:C04
	M00001360A:G10	M00001641C:H07
	M00004216D:C03	M00001641C:F01
	M00004076D:D04	M00001641C:D02
	M00001484C:A04	M00001641B:F12
	M00001456B:G01	M00001634A:B04
	M00003972D:C09	M00001636B:G11
	M00003974C:E04	M00001649C:D05

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES29 ATCC No. 207076	cDNA Library Ref ES30 ATCC No. 207077
	M00003979A:E11	M00001636A:C03
	M00003983C:F03	M00001635D:D05
	M00003989B:F11	M00001635D:C12
	M00004031D:B05	M00001635B:H02
	M00004177C:A01	M00001635B:H01
	M00004076B:G03	M00001634D:G11
	M00004167D:A07	M00001634D:D04
	M00004078A:A06	M00001634A:H05
	M00004085A:B02	M00001641A:A11
	M00004107B:A06	M00001638B:E12
	M00004111C:E11	M00001640A:H02
	M00004130D:H01	M00001614C:E06
	M00004157D:B03	M00001636D:F09
	M00004159C:F09	M00001637A:A03
	M00004162C:A07	M00001637A:A06
	M00004135B:G01	M00001637A:E10
	M00004040A:G12	M00001637A:F10
	M00001453B:H12	M00001637C:C06
	M00001448A:E11	M00001644A:H01
	M00001448B:F09	M00001638B:E03
	M00001448B:H05	M00001649A:E11
	M00001448C:E11	M00001638B:F10
	M00001448C:F10	M00001639A:C03
	M00001448D:F12	M00001639A:G07
	M00001449B:B03	M00001639B:H01
	M00001449C:C05	M00001639B:H05
	M00001449D:G10	M00001639C:A09
	M00001448A:B12	M00001639C:C02
	M00001453A:D08	M00001649C:E11
	M00001451B:A04	M00001649C:H10
	M00001454A:F11	M00001637C:E03
	M00001454A:G03	M00001617A:A08
	M00001455A:F04	M00001622A:H12
	M00001455B:E07	M00001621C:H12
	M00001455D:A06	M00001621B:G05
	M00001364B:B06	M00001620D:H02
	M00004117A:G01	M00001620D:G11
	M00001455D:D11	M00001619D:D10
	M00001456B:A06	M00001619C:C07
	M00001451A:C10	M00001619A:E05
	M00001395A:E03	M00001623A:F04
	M00001366D:C06	M00001618A:A03
	M00001365A:H10	M00001618B:D09
	M00001366D:C12	M00001617A:A01
	M00001373D:B03	M00001616D:C11
	M00001453B:F08	M00001615C:G05
	M00001444D:C01	M00001615C:A11
	M00001375B:C06	M00001615B:G07

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES29 ATCC No. 207076	cDNA Library Ref ES30 ATCC No. 207077
	M00001392C:D05	M00001633D:H06
	M00001395A:A12	M00001639C:A10
	M00001395A:H02	M00001615B:A09
	M00001397D:G08	M00001615B:G01
	M00001434A:B10	M00001618A:F10
	M00001416A:D09	M00001632C:H07
	M00001433C:F10	M00001633D:D12
	M00001416A:H02	M00001633D:D09
	M00001428D:B10	M00001618A:F08
	M00001428B:D01	M00001633D:G09
	M00001426D:D12	M00001624A:A03
	M00001400C:D02	M00001633C:F09
	M00001427C:D01	M00001633C:H05
		M00001633C:B09
		M00001633A:E06
		M00001633C:H11
		M00001632C:B10
		M00001625D:G10
		M00001631D:G05
		M00001629C:E07
		M00001629B:B08
		M00001626C:E04
		M00001626C:C11
		M00001632A:B10
		M00001624B:B10
		M00001633C:A05
		M00001625C:G05

Table 40. Clones Deposited on January 22, 1999

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES31 ATCC No. 207078	cDNA Ref No. ES32 ATCC No. 207079	cDNA Library Ref ES33 ATCC No. 207080
Clone Names in Library	M00003843A:E04	M00003906A:F12	M00005254D:A10
	M00003842C:G03	M00003906B:H06	M00005260B:E11
	M00003842A:A03	M00003906C:C05	M00005260A:F04
	M00003841D:A04	M00003907A:F01	M00005260A:A12
	M00003841B:E06	M00003907B:C03	M00005259B:D12
	M00003841C:H11	M00003907B:D05	M00005257D:H11
	M00003844A:A11	M00003918A:D08	M00005257D:G07
	M00003841C:F01	M00003918A:F09	M00005257D:A06
	M00003841C:H08	M00003918C:H10	M00005257C:G01
	M00003841C:D07	M00003924A:D08	M00005257A:H11
	M00003844D:A07	M00003958B:E11	M00005236B:H10
	M00003845D:G08	M00003958B:H08	M00005236B:G03
	M00003852C:B06	M00003960A:G07	M00005257C:E05
	M00003854B:A07	M00003971B:A10	M00001608C:D02
	M00003854B:D04	M00003972D:H02	M00001608C:G04
	M00003859D:C05	M00003973C:C03	M00001608D:F11

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES31 ATCC No. 207078	cDNA Ref No. ES32 ATCC No. 207079	cDNA Library Ref ES33 ATCC No. 207080
	M00003860B:F11	M00003974B:B11	M00001609C:A12
	M00003867B:G07	M00003974D:F02	M00001609C:G05
	M00003867B:G08	M00003974D:H04	M00001610C:B07
	M00003841B:E03	M00003975C:F07	M00001612D:D12
	M00003822D:B10	M00003977C:A06	M00001612D:F06
	M00003867D:A06	M00003977C:B03	M00001613A:D02
	M00003868B:G06	M00003977D:A03	M00001614A:B10
	M00003867B:D10	M00003977D:A06	M00001614C:G07
	M00003831C:G05	M00003977D:D04	M00001615C:E07
	M00003901C:B01	M00003978D:G04	M00001625C:F10
	M00003868C:C07	M00003980A:F04	M00001626D:A02
	M00003820A:A08	M00003980B:C11	M00001629A:H09
	M00003820B:D07	M00003981C:B04	M00001629D:B10
	M00003820B:D10	M00003982A:B12	M00001629D:D10
	M00003822D:C06	M00003982C:G04	M00001630C:F09
	M00003823B:F07	M00003984D:B08	M00001631A:D03
	M00003824C:D07	M00003985B:G04	M00001631A:F06
	M00003825B:B10	M00003985D:E10	M00001631A:F12
	M00003825B:B11	M00003986B:A08	M00001631B:H04
	M00003828A:D05	M00003986C:D09	M00001633A:F11
	M00003822D:D04	M00003986D:C08	M00001633A:G10
	M00003830C:A03	M00003987B:E12	M00001633B:A12
	M00003840D:H10	M00003987B:F08	M00001633B:E03
	M00003832A:A09	M00003987C:G03	M00001633C:A08
	M00003833B:B03	M00003988D:A08	M00001633C:E12
	M00003833B:C12	M00003989C:D03	M00001635B:B02
	M00003834B:G04	M00003989C:G05	M00001636A:H12
	M00003835A:A09	M00003989D:F12	M00001638A:C08
	M00003835B:H11	M00004029B:F01	M00001638B:C08
	M00003835D:G06	M00004029C:C05	M00001639D:C12
	M00003837C:E05	M00004029C:G10	M00001640A:F05
	M00003837C:F10	M00004030D:F11	M00001642D:G08
	M00003839A:D07	M00004034A:A01	M00001647D:G07
	M00003839D:E11	M00004034C:G02	M00001649A:E10
	M00003829C:H05	M00004034D:E09	M00001650D:D10
	M00003901B:C03	M00004035B:H09	M00001650D:F11
	M00003878C:F06	M00004036D:B04	M00001651C:D11
	M00003878C:G08	M00004036D:B09	M00001651C:G12
	M00003879A:A02	M00004038A:F02	M00001652B:D06
	M00003879A:B08	M00004038D:G06	M00001652D:G02
	M00003879A:C11	M00004039A:C03	M00001652D:G06
	M00003879A:D02	M00004039A:H11	M00001653A:A05
	M00003879B:G02	M00004039B:A05	M00001653D:H07
	M00003880B:D11	M00004039B:E12	M00001654A:E08
	M00003880C:E11	M00004040C:A01	M00001654B:A01
	M00003880C:H03	M00004051D:E01	M00001654C:D10
	M00003901B:F10	M00004072D:F09	M00001654C:G07
	M00003890B:C08	M00004073A:D10	M00001654C:G09

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES31 ATCC No. 207078	cDNA Ref No. ES32 ATCC No. 207079	cDNA Library Ref ES33 ATCC No. 207080
	M00003877C:A11	M00004075B:G09	M00001655C:C07
	M00003819D:B01	M00004076A:D12	M00001655D:E08
	M00003901B:G11	M00004076D:H07	M00001655D:H11
	M00001692A:G06	M00004078A:C11	M00001656A:H12
	M00003903C:C05	M00004078A:E05	M00001656C:C04
	M00003903C:E12	M00004078A:F07	M00001656D:C04
	M00003903D:C12	M00004078B:C11	M00001657C:C11
	M00003903D:D10	M00004078B:F12	M00001657D:A10
	M00003903D:H11	M00004079D:G08	M00001659D:A09
	M00003904A:C04	M00004081A:E02	M00001661D:D05
	M00003904B:C03	M00004081A:G01	M00001664B:E08
	M00003904C:A08	M00004081C:A10	M00001664B:F06
	M00003881B:F10	M00004083A:E08	M00001669B:C12
	M00003871D:G06	M00004083B:C01	M00001669C:B09
	M00003868D:D09	M00004086D:G08	M00001670A:F09
	M00003868D:D11	M00004087B:A12	M00001678C:F09
	M00003870C:A01	M00004087C:A01	M00001693A:H06
	M00003870C:A10	M00004088C:F01	M00003805D:E06
	M00003870C:E10	M00004088D:A11	M00003806C:A06
	M00003871A:A02	M00004088D:B05	M00003809B:A03
	M00003871A:B09	M00004088D:B10	M00003810A:A02
	M00003871A:C11	M00004090B:B04	M00003810B:B11
	M00003871A:G09	M00004090B:H06	M00003810C:B06
	M00003871C:E04	M00004092B:E05	M00003810D:H09
	M00003871C:F12	M00004093C:C02	M00003811C:C02
	M00003878C:D08	M00004096D:H03	M00003813B:F02
	M00003871D:E11	M00004099D:F01	M00003813C:H08
	M00003877C:G12	M00004100B:C07	M00003813D:B12
	M00003875A:A07	M00004103B:E09	M00003813D:C02
	M00003875A:B01	M00004105C:B05	M00003813D:G06
	M00003875B:F12	M00004105C:C08	M00003814B:C01
	M00003875C:A01	M00004107A:A12	M00003817C:A10
	M00003875C:A09	M00004107B:D07	M00003817C:G06
	M00003875C:G02	M00004108B:B02	M00003817D:D12
	M00003876B:C05	M00004108D:E07	M00003821A:H09
	M00003876C:D02	M00004108D:G04	M00003822B:G12
	M00003876C:F02	M00004110A:A10	M00003822C:A07
	M00003877B:H10	M00004110B:A07	M00003823C:B01
	M00003868D:B09	M00004118B:A03	M00003823C:C04
	M00003871D:A10	M00004118B:F01	M00003824A:G11
	M00001669D:D06	M00004118D:B05	M00003824B:C09
	M00001661A:B11	M00004119A:C09	M00003824C:A10
	M00001661B:F06	M00004136D:B02	M00003824D:D08
	M00001662A:C07	M00004137A:D06	M00003825B:F10
	M00001662A:G01	M00004139C:A12	M00003825D:F01
	M00001662B:F06	M00004149C:B02	M00003826C:F05
	M00001663C:F12	M00004159C:G12	M00003829A:B08
	M00001664A:F08	M00004169D:B11	M00003829C:E08

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES31 ATCC No. 207078	cDNA Ref No. ES32 ATCC No. 207079	cDNA Library Ref ES33 ATCC No. 207080
	M00001664D:F04	M00004187D:H06	M00003829D:D12
	M00001661A:E06	M00004228C:H03	M00003829D:F03
	M00001669A:B02	M00004244C:G07	M00003830D:B11
	M00001669B:B12	M00004358D:C02	M00003830D:H11
	M00001669C:C08	M00004690A:G08	M00003833D:H08
	M00001675A:G10	M00004891B:D01	M00003833D:H10
	M00001669D:C03	M00004891C:D04	M00003840A:C10
	M00001660B:E03	M00004895B:E12	M00003840B:F05
	M00001669D:F05	M00004895B:G04	M00003840C:C02
	M00001670B:G12	M00004895D:G07	M00003845C:D04
	M00001671A:A10	M00004898C:F03	M00003845D:A04
	M00001671B:G05	M00004899D:G06	M00003846B:C05
	M00001671C:C11	M00004959D:H12	M00003846C:F08
	M00001672D:E08	M00004960A:B08	M00003848B:E07
	M00001673A:G08	M00004960C:E10	M00003848D:G02
	M00001673B:B07	M00005100A:B02	M00003850C:G09
	M00001673B:F07	M00005100A:C01	M00003851A:A06
	M00001673D:D06	M00005101C:E12	M00003851B:D03
	M00001673D:F10	M00005102C:D03	M00003851B:E01
	M00001674A:G07	M00005134B:E08	M00003851C:F09
	M00001692D:B01	M00005139A:H03	M00003851D:H11
	M00001669C:D09	M00005140C:B10	M00003852B:G04
	M00001655C:E01	M00005140D:C06	M00003852C:F07
	M00001649D:A08	M00005178D:H04	M00003853B:C10
	M00001650A:C11	M00005210A:E06	M00003854C:C09
	M00001651A:H11	M00005212B:E01	M00003855A:A01
	M00001652A:A01	M00005212C:C03	M00003855A:F01
	M00001652B:G10	M00005212C:D02	M00003855B:B09
	M00001652D:E05	M00005212C:H02	M00003856A:G04
	M00001652D:E09	M00005212D:D09	M00003856B:A12
	M00001653B:C06	M00005212D:H01	M00003857A:E12
	M00001653B:G10	M00005216A:D09	M00003857A:H10
	M00001653C:D10	M00005216A:H01	M00003857C:E05
	M00001654D:A03	M00005217B:A06	M00003858B:G02
	M00001654D:E12	M00005218A:F09	M00003860D:E06
	M00001654D:F11	M00005228A:B03	M00003905C:F12
	M00001660C:B06	M00005228C:C05	M00003911A:D12
	M00001658D:G12	M00005229B:G12	M00003966B:A04
	M00001675C:A04	M00005229B:H04	M00003966C:A12
	M00001660B:D03	M00005229B:H06	M00003966C:F03
	M00001660B:A09	M00005229D:H03	M00003973D:F08
	M00001659D:C09	M00005230B:H09	M00003974D:E01
	M00001659D:B05	M00005232A:H12	M00003974D:H07
	M00001654D:F12	M00005233B:D04	M00003976B:E06
	M00001659A:D12	M00005233D:H07	M00003976B:H07
	M00001655A:B11	M00005235B:F10	M00003978A:E01
	M00001658B:A07	M00005236A:E04	M00003978A:E09
	M00001658A:G09	M00005236A:G10	M00003978C:A12

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES31 ATCC No. 207078	cDNA Ref No. ES32 ATCC No. 207079	cDNA Library Ref ES33 ATCC No. 207080
	M00001657D:A04	M00005236B:A12	M00003980C:E12
	M00001657B:B04	M00001448B:A07	M00003980C:F12
	M00001656B:E01	M00001448B:G07	M00003981A:A07
	M00001660B:E04	M00001448D:E11	M00003981B:B12
	M00001659C:F10	M00001455A:D10	M00003982A:G03
	M00003808C:A05	M00001455A:E11	M00003982B:C10
	M00001694D:C12	M00001476D:F12	M00003982B:H10
	M00003746C:E02	M00001478A:F12	M00003983A:D02
	M00003779D:E08	M00001482C:F09	M00003983A:F06
	M00003792A:B10	M00001485C:D07	M00003983A:G02
	M00003793D:A11	M00001485C:G06	M00003983D:E08
	M00003794D:G03	M00001485D:A05	M00003983D:H02
	M00003797A:C11	M00001487C:A11	M00003985A:C01
	M00003797A:D06	M00001487C:G09	M00003986C:G11
	M00003797A:G03	M00001530A:B02	M00003986D:H12
	M00003800B:F03	M00001530A:H05	M00004027A:A08
	M00003805A:F02	M00001530D:A11	M00004028A:B10
	M00003806B:C09	M00001539B:B10	M00004028A:G03
	M00001674A:G11	M00001567A:C04	M00004029B:A01
	M00003806D:D11	M00001567A:C11	M00004029B:A06
	M00001693D:E08	M00001567C:B08	M00004029B:G10
	M00003808D:D08	M00001567C:E07	M00004029C:F02
	M00003809A:C01	M00001570C:B02	M00004029C:F05
	M00003809A:F01	M00001570D:E05	M00004030B:A12
	M00003809B:B02	M00001570D:E07	M00004030B:D08
	M00003809B:E10	M00001573B:A06	M00004030C:A08
	M00003813A:B02	M00001573B:H12	M00004030C:C02
	M00003813A:D08	M00001575A:D05	M00004034C:F05
	M00003813B:E09	M00001575B:C01	M00004035B:F05
	M00003814B:C12	M00001576C:H02	M00004036A:A11
	M00003814B:F12	M00001577A:A03	M00004037C:D04
	M00003815C:C06	M00001578B:A06	M00004038A:E05
	M00003815C:D12	M00001579D:F02	M00004038B:D01
	M00003817B:C04	M00001582C:C04	M00004039C:E02
	M00003806B:G05	M00001582C:G02	M00004039D:B10
	M00001679A:D10	M00001584A:A07	M00004040A:A07
	M00001675C:C03	M00001584D:B06	M00004040A:B04
	M00001675C:D12	M00001584D:C11	M00004040A:C08
	M00001675D:E10	M00001585D:B12	M00004040B:C05
	M00001676B:B09	M00001586C:H07	M00004040B:F07
	M00001676B:E01	M00001589D:A01	M00004069A:E12
	M00001676C:A04	M00001590D:B04	M00004069C:C08
	M00001676C:E07	M00001592B:B02	M00004077A:G12
	M00001676D:A02	M00001592D:H02	M00004085B:G01
	M00001676D:B02	M00001594C:E05	M00004087A:B05
	M00001677A:G11	M00001594C:H03	M00004090D:F12
	M00001677B:A12	M00001594D:G11	M00004092C:D08
	M00001677B:B04	M00001595A:C07	M00004097C:E03

cDNA Ref No.; ATCC Accession No.	cDNA Library Ref ES31 ATCC No. 207078	cDNA Ref No. ES32 ATCC No. 207079	cDNA Library Ref ES33 ATCC No. 207080
	M00001677D:B01	M00001595A:D12	M00004097C:H08
	M00001678D:B11	M00001595A:E07	M00004097D:B05
	M00001681C:A08	M00001595B:G07	
	M00003819B:G01	M00001595B:G10	
	M00001693C:E09	M00001595B:H11	
	M00001693C:C12	M00001595C:A01	
	M00001692B:E01	M00001595C:A05	
	M00001692A:B06	M00001595C:B12	
	M00001678B:H01	M00001595C:E05	
	M00001681D:C12	M00001595C:E09	
	M00001694A:E03	M00001595D:C11	
	M00001680B:D02	M00001596A:A02	
	M00001680A:B02	M00001596A:D01	
	M00001679D:F02	M00001596C:G05	
	M00001679D:B02	M00001607A:A01	
	M00001679A:G06		

Retrieval of Individual Clones from Deposit of Pooled Clones

Where the ATCC deposit is composed of a pool of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones were then deposited as a pool of equal mixtures in the composite deposit. Particular clones can be obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a T_m of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C). Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

Example 27: Source of Biological Materials and Overview of Novel Polynucleotides Expressed by the Biological Materials

cDNA libraries were constructed from either human colon cancer cell line Km12L4-A (Morikawa, et al., *Cancer Research* (1988) *48*:6863) , KM12C (Morikawa et al. *Cancer Res.* (1988) *48*:1943-1948), or MDA-MB-231 (Brinkley et al. *Cancer Res.* (1980) *40*:3118-3129) was used to construct a cDNA library from mRNA isolated from the cells. Sequences expressed by these cell lines were isolated and analyzed; most sequences were about 275-300 nucleotides in length. The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B_2 surgical specimen (Morikawa *et al. Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman *et al. Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling *et al. Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as a model cell line for the study of colon cancer (see, *e.g.*, Moriakawa *et al.*, *supra*; Radinsky *et al. Clin. Cancer Res.* (1995) *1*:19; Yeatman *et al.*, (1995) *supra*; Yeatman *et al. Clin. Exp. Metastasis* (1996) *14*:246). The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma.

The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol.* 266:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis Techniques" Adams *et al.*, eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie *et al. Comput. Chem.* (1993) 17:191). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats. Masking resulted in the elimination of 43 sequences. The remaining sequences were then used in a BLASTN vs. GenBank search; sequences that exhibited greater than 70% overlap, 99% identity, and a p value of less than 1×10^{-40} were discarded. Sequences from this search also were discarded if the inclusive

parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database search: (1) unknown (no hits in the GenBank search), (2) weak similarity (greater than 45% identity and p value of less than 1×10^{-5}), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1×10^{-5}). Sequences having greater than 70% overlap, greater than 99% identity, and p value of less than 1×10^{-40} were discarded.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search was performed and sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1×10^{-40} were discarded. Sequences with a p value of less than 1×10^{-65} when compared to a database sequence of human origin were also excluded. Second, a BLASTN vs. Patent GeneSeq database was performed and sequences having greater than 99% identity, p value less than 1×10^{-40}, and greater than 99% overlap were discarded.

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1×10^{-111} in relation to a database sequence of human origin were specifically excluded. The final result provided the 1,565 sequences listed as SEQ ID NOS:6097-7661 in the accompanying Sequence Listing and summarized in Table 41A (inserted prior to claims). Each identified polynucleotide represents sequence from at least a partial mRNA transcript.

Table 41A provides: 1) the SEQ ID NO assigned to each sequence for use in the present specification; 2) the filing date of the U.S. priority application in which the sequence was first filed; 3) the attorney docket number assigned to the priority application (for internal use); 4) the SEQ ID NO assigned to the sequence in the priority application; 5) the sequence name used as an internal identifier of the sequence; and 6) the name assigned to the clone from which the sequence was isolated. Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOS: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene.

In order to confirm the sequences of SEQ ID NOS: 6097-7661, the clones were retrieved from a library using a robotic retrieval system, and the inserts of the retrieved clones re-sequenced. These "validation" sequences are provided as SEQ ID NOS:7662-8706 in the Sequence Listing, and a summary of the "validation" sequences provided in Table 41B (inserted prior to claims). Table 41B provides: 1) the SEQ ID NO assigned to each sequence for use in the present specification; 2) the sequence name assigned to the "validation" sequence obtained; 3) whether the "validation" sequence contains sequence that overlaps with an original sequence of SEQ ID NOS: 6097-7661 (Validation Overlap (VO)), or whether the "validation" sequence does not substantially overlap with an original sequence of SEQ ID NOS: 6097-7661 (indicated by Validation Non-Overlap (VNO)); and 4) where the sequence is indicated as VO, the name of the clone that contains the indicated "validation" sequence. "Validation" sequences are indicated as "VO" where the "validation" sequence overlaps with an original sequence (*e.g.*, one of SEQ ID NOS: 6097-7661), and/or the "validation" sequence belongs to the same cluster as the original sequence using the clustering technique described above. Because the inserts of the clones are generally longer than the original sequence and the validation sequence, it is possible that a "validation" sequence can be obtained from the same clone as an original sequence but yet not share any of the sequence of the original. Such validation sequences will, however, belong to the same cluster as the original sequence using the clustering technique described above. VO "validation" sequences are contained within the same clone as the original sequence (one of SEQ ID NOS: 6097-7661). "Validation" sequences that provided overlapping sequence are indicating by "VO" can be correlated with the original sequences they validate by referring to Table 41A. Sequences indicated as VNO are treated as newly isolated sequences and may or may not be related to the sequences of SEQ ID NOS: 6097-7661. Because the "validation" sequences are often longer than the original polynucleotide sequences and thus provide additional sequence information. All validation sequences can be obtained either from an indicated clone (*e.g.*, for VO sequences) or from a cDNA library described herein (*e.g.*, using primers designed from the sequence provided in the sequence listing).

Example 28: Results of Public Database Search to Identify Function of Gene Products

SEQ ID NOS: 7662-8706 were translated in all three reading frames, and the nucleotide sequences and translated amino acid sequences used as query sequences to

search for homologous sequences in either the GenBank (nucleotide sequences) or Non-Redundant Protein (amino acid sequences) databases. Query and individual sequences were aligned using the BLAST 2.0 programs, available over the world wide web of the NCBI. (see also Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402). The sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above.

Tables 41A and 41B (inserted before the claims) provide the alignment summaries having a p value of 1×10^{-2} or less indicating substantial homology between the sequences of the present invention and those of the indicated public databases. Table 41A provides the SEQ ID NO of the query sequence, the accession number of the GenBank database entry of the homologous sequence, and the p value of the alignment. Table 41A provides the SEQ ID NO of the query sequence, the accession number of the Non-Redundant Protein database entry of the homologous sequence, and the p value of the alignment. The alignments provided in Tables 41A and 41B are the best available alignment to a DNA or amino acid sequence at a time just prior to filing of the present specification. The activity of the polypeptide encoded by the SEQ ID NOS listed in Tables 41A and 41B can be extrapolated to be substantially the same or substantially similar to the activity of the reported nearest neighbor or closely related sequence. The accession number of the nearest neighbor is reported, providing a publicly available reference to the activities and functions exhibited by the nearest neighbor. The public information regarding the activities and functions of each of the nearest neighbor sequences is incorporated by reference in this application. Also incorporated by reference is all publicly available information regarding the sequence, as well as the putative and actual activities and functions of the nearest neighbor sequences listed in Table 41 and their related sequences. The search program and database used for the alignment, as well as the calculation of the p value are also indicated.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence of the corresponding polynucleotide. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of the corresponding polynucleotides.

Table 41A: Nearest Neighbor (BlastN vs. Genbank)

SEQ ID	ACC'N	DESCRIP.	P VALUE
6667	L17043	Homo sapiens pregnancy-specific beta-1-glycoprotein-11 gene.	1.00E-12
6674	M18864	Rat bone protein I (BP-I) mRNA, partial cds.	7.00E-30
6705	L13838	Human genomic sequence from chromosome 13, clone ch13lambdacDNA17-18.	4.00E-36
6714	U09646	Human carnitine palmitoyltransferase II precursor	1.00E-34
6723	U72621	Human LOT1 mRNA, complete cds	1.00E-43
6725	M20910	Human 7S L gene, complete.	1.00E-35
6732	Z48950	H.sapiens hH3.3B gene for histone H3.3	4.00E-36
6735	X00247	Human translocated c-myc gene in Raji Burkitt lymphoma cells	3.00E-44
6739	D80007	Human mRNA for KIAA0185 gene, partial cds	7.00E-52
6742	U14967	Human ribosomal protein L21 mRNA, complete cds.	2.00E-42
6745	M13934	Human ribosomal protein S14 gene, complete cds.	4.00E-45
6748	NM_003902.1	Homo sapiens far upstream element binding protein (FUBP) mRNA > :: gb|U05040|HSU05040 Human FUSE binding protein mRNA, complete cds.	1.00E-54
6753	L41142	Homo sapiens signal transducer and activator of transcription (STAT5) mRNA, complete cds.	2.00E-62
6761	Z12112	pWE15A cosmid vector DNA	2.00E-52
6763	Z54386	H.sapiens CpG island DNA genomic Mse1 fragment, clone 10g3, forward read cpg10g3.ft1a	7.00E-48
6764	X80333	M.musculus rab18 mRNA	2.00E-52
6765	X52126	Human alternatively spliced c-myb mRNA	1.00E-64
6767	L26247	Homo sapiens sui1iso1 mRNA, complete cds.	3.00E-54
6772	NM_001736.1	Homo sapiens complement component 5 receptor 1 C5a anaphylatoxin receptor mRNA, complete cds.	4.00E-56
6773	Z50798	G.gallus mRNA for p52	4.00E-55
6775	AB002368	Human mRNA for KIAA0370 gene, partial cds	2.00E-58
6777	M26697	Human nucleolar protein (B23) mRNA, complete cds.	4.00E-48

6779	D42087	Human mRNA for KIAA0118 gene, partial cds	4.00E-56
6789	D50734	Rat mRNA of antizyme inhibitor, complete cds	2.00E-50
6793	X02344	Homo sapiens beta 2 gene	1.00E-67
6794	NM_001067.1	Homo sapiens topoisomerase (DNA) II alpha topoisomerase II (top2) mRNA, complete cds.	7.00E-63
6797	U36309	Gallus gallus rhoGap protein mRNA, complete cds	3.00E-62
6799	NM_002842.1	Homo sapiens protein tyrosine phosphatase, receptor type, H (PTPRH) mRNA > :: dbj|D15049|HUMSAP1C Human mRNA for protein tyrosine phosphatase	2.00E-81
6803	U47322	Cloning vector DNA, complete sequence.	1.00E-63
6810	NM_001190.1	Homo sapiens branched chain aminotransferase 2, mitochondrial (BCAT2) mRNA > :: gb|U68418|HSU68418 Human branched chain aminotransferase precursor (BCATm) mRNA, nuclear gene encoding mitochondrial protein, complete cds	4.00E-67
6814	S62077	HP1Hs alpha=25 kda chromosomal autoantigen [human, mRNA, 876 nt]	5.00E-68
6815	U34991	Human endogenous retrovirus clone c18.4, HERV-H/HERV-E hybrid multiply spliced protease/integrase mRNA, complete cds, and envelope protein mRNA, partial cds	2.00E-61
6818	U18671	Human Stat2 gene, complete cds.	4.00E-77
6819	L18964	Human protein kinase C iota isoform (PRKCI) mRNA, complete cds.	4.00E-68
6820	D29956	Human mRNA for KIAA0055 gene, complete cds	6.00E-70
6821	M77140	H.sapiens pro-galanin mRNA, 3' end.	2.00E-72
6824	U51432	Homo sapiens nuclear protein Skip mRNA, complete cds	1.00E-75
6825	M84334	Macacca mulatta hnRNP A1-gamma isoform mRNA, complete cds.	5.00E-50

6826	NM_002592.1	Homo sapiens proliferating cell nuclear antigen (PCNA) mRNA > :: gb|M15796|HUMCYL Human cyclin protein gene, complete cds.	1.00E-74
6827	M88458	Human ELP-1 mRNA sequence.	4.00E-76
6828	U44940	Mus musculus quaking type I (QKI) mRNA, complete cds	2.00E-69
6829	D17577	Mouse mRNA for kinesin-like protein (Kif1b), complete cds	2.00E-71
6830	U18920	Human chromosome 17q12-21 mRNA, clone pOV-3, partial cds.	2.00E-72
6832	M21188	Human insulin-degrading enzyme (IDE) mRNA, complete cds.	7.00E-82
6833	U49058	Rattus norvegicus CTD-binding SR-like protein rA4 mRNA, partial cds	1.00E-67
6835	D10630	Mus musculus mRNA for zinc finger protein, complete cds, clone:CTfin51	4.00E-76
6836	U29156	Mus musculus eps15R mRNA, complete cds.	3.00E-84
6837	Y08135	M.musculus mRNA for ASM-like phosphodiesterase 3a	1.00E-86
6838	U90567	Gallus gallus glutamine rich protein mRNA, partial cds	5.00E-58
6839	U58280	Mus musculus second largest subunit of RNA polymerase I (RPA2) mRNA, complete cds	4.00E-77
6840	S79539	Pat-12=Pat-12 product [mice, embryonic stem ES cells, mRNA, 2781 nt]	9.00E-84
6841	D30666	Rat mRNA for brain acyl-CoA synthetase II, complete cds	2.00E-89
6842	U29156	Mus musculus eps15R mRNA, complete cds.	2.00E-92
6844	U36909	Bos taurus Rho-associated kinase mRNA, complete cds	e-104
6845	L36315	Mus musculus (clone pMLZ-1) zinc finger protein	e-105
6846	X80169	M.musculus mRNA for 200 kD protein	e-106

6847	X83577	M.musculus mRNA for K-glypican	e-107
7156	Z95437	Human DNA sequence from cosmid A1 on chromosome 6 contains ESTs. HERV like retroviral sequence	8.00E-21
7208	X69907	H.sapiens gene for mitochondrial ATP synthase c subunit (P1 form)	6.00E-07
7221	M19390	Bovine interstitial retinol binding protein	8.00E-31
7252	U19247	Homo sapiens interferon-gamma receptor alpha chain gene, exon 7 and complete cds	7.00E-41
7266	U20239	Mus musculus fibrosin mRNA, partial cds	5.00E-38
7267	D26361	Human mRNA for KIAA0042 gene, complete cds	2.00E-41
7291	NM_000694.1	Homo sapiens aldehyde dehydrogenase 7 (ALDH7) mRNA > :: gb|U10868|HSU10868 Human aldehyde dehydrogenase ALDH7 mRNA, complete cds.	1.00E-37
7292	U84404	Human E6-associated protein E6-AP/ubiquitin-protein ligase (UBE3A) mRNA, alternatively spliced, complete cds	1.00E-46
7299	U51714	Human GPI protein p137 mRNA, partial sequence, 3'-UTR.	9.00E-53
7300	U58884	Mus musculus SH3-containing protein SH3P7 mRNA, complete cds. similar to Human Drebrin	2.00E-49
7306	X79067	H.sapiens ERF-1 mRNA 3' end	2.00E-72
7308	U00946	Human clone A9A2BRB5 (CAC)n/(GTG)n repeat-containing mRNA	3.00E-54
7313	D11078	Homo sapiens RGH2 gene, retrovirus-like element	6.00E-49
7315	U05989	Rattus norvegicus clone par-4 induced by effectors of apoptosis mRNA, complete cds.	3.00E-64
7316	U13185	Cloning vector pbetagal-Enhancer, complete sequence.	3.00E-52
7318	D87443	Human mRNA for KIAA0254 gene, complete cds	8.00E-63

7321	U19867	Cloning vector pSPL3, exon splicing vector, complete sequence, HIV envelope protein gp160 and beta-lactamase, complete cds.	7.00E-72
7323	U04817	Human protein kinase PITSLRE alpha 2-3 mRNA, complete cds.	4.00E-57
7326	U03687	Photinus pyralis modified luciferase gene, complete cds, and pUC18 derived vector.	3.00E-62
7327	U27196	Gallus gallus zinc finger protein (Fzf-1) mRNA, complete cds.	1.00E-66
7331	X53586	Human mRNA for integrin alpha 6	2.00E-71
7332	J05016	Human (clone pA3) protein disulfide isomerase related protein (ERp72) mRNA, complete cds.	3.00E-67
7333	M86752	Human transformation-sensitive protein (IEF SSP 3521) mRNA, complete cds.	1.00E-66
7335	L19437	Human transaldolase mRNA containing transposable element, complete cds	5.00E-70
7337	X90857	H.sapiens mRNA for -14 gene, containing globin regulatory element	1.00E-74
7338	NM_003980.1	Homo sapiens microtubule associated protein 7 mRNA	9.00E-76
7341	U17901	Rattus norvegicus phospholipase A-2-activating protein (plap) mRNA, complete cds.	3.00E-75
7342	S80632	threonine, tyrosine phosphatase [human, brain, mRNA Partial, 1236 nt]	2.00E-69
7343	M76541	Human DNA-binding protein (NF-E1) mRNA, complete cds.	2.00E-80
7344	S55305	14-3-3 protein gamma subtype=putative protein kinase C regulatory protein [rats, brain, mRNA, 3410 nt] > :: dbj|D17447|D17447 Rattus norvegicus mRNA for 14-3-3 protein gamma-subtype, complete cds	7.00E-93

7345	NM_002350.1	Homo sapiens v-yes-1 Yamaguchi sarcoma viral related oncogene homolog (LYN) mRNA > :: gb|M16038|HUMLYN Human lyn mRNA encoding a tyrosine kinase.	3.00E-86
7346	Y10725	M.musculus mRNA for protein kinase KIS	4.00E-68
7347	U89931	Cloning vector pTRE, complete sequence	3.00E-65
7348	Z46386	Bovine herpesvirus type 4 DNA for nonconserved region F (DN599 like strain)	3.00E-73
7349	L77599	Homo sapiens (clone SEL214) 17q YAC (303G8) RNA.	2.00E-69
7351	Y10746	H.sapiens mRNA for protein containing MBD 1	2.00E-79
7352	L77599	Homo sapiens (clone SEL214) 17q YAC (303G8) RNA.	2.00E-71
7353	Z57619	H.sapiens CpG island DNA genomic Mse1 fragment, clone 187a6, forward read cpg187a6.ft1b	7.00E-72
7354	U48807	Human MAP kinase phosphatase (MKP-2) mRNA, complete cds	3.00E-76
7356	M27444	Bos taurus (clone pTKD7) dopamine and cyclic AMP-regulated neuronal phosphoprotein (DARPP-32) mRNA, complete cds.	4.00E-76
7357	U37150	Bos taurus peptide methionine sulfoxide reductase (msrA) mRNA, complete cds	5.00E-78
7358	U02435	Cloning vector pSVbeta, complete sequence	1.00E-77
7359	U09662	Cloning vector pSEAP-Enhancer, complete sequence	4.00E-79
7360	M99566	sCos cloning vector SfiI containing bacteriophage promoters and flanking restriction sites in sCos vectors.	1.00E-79
7362	Z12112	pWE15A cosmid vector DNA	4.00E-80
7363	U55387	Cricetulus griseus SL15 mRNA, complete cds	2.00E-82
7365	L14684	Rattus norvegicus nuclear-encoded mitochondrial elongation factor G mRNA, complete cds.	2.00E-91

7366	U49057	Rattus norvegicus CTD-binding SR-like protein rA9 mRNA, complete cds	7.00E-93
7367	U57368	Mus musculus EGF repeat transmembrane protein mRNA, complete cds.	4.00E-97
7368	AF000938	Mus musculus RNA polymerase I largest subunit	8.00E-94
7370	X80169	M.musculus mRNA for 200 kD protein	e-102
7371	U09874	Mus musculus SKD3 mRNA, complete cds.	e-105
7372	D78020	Rat mRNA for NFI-A4, partial cds	e-108
7611	Z73360	Human DNA sequence from cosmid 92M18, BRCA2 gene region chromosome 13q12-13	9.00E-22
7618	X62078	H.sapiens mRNA for GM2 activator protein	7.00E-72
7619	X85750	H.sapiens mRNA for transcript associated with monocyte to macrophage differentiation	2.00E-50
7621	X03473	Human gene for histone H1(0)	1.00E-67
7631	X64411	R.norvegicus mRNA for 100 kDa protein	1.00E-54
7634	X13345	Human gene for plasminogen activator inhibitor 1	2.00E-59
7638	D86971	Human mRNA for KIAA0217 gene, partial cds	7.00E-83
7639	NM_001859.1	Homo sapiens solute carrier family 31 gb|U83460|HSU83460 Human high-affinity copper uptake protein (hCTR1) mRNA, complete cds	7.00E-72
7640	X68194	H.sapiens h-Sp1 mRNA	5.00E-57
7641	AB002326	Human mRNA for KIAA0328 gene, partial cds	3.00E-74
7644	D31762	Human mRNA for KIAA0057 gene, complete cds	3.00E-85
7646	X58472	Mouse KIN17 mRNA for kin17 protein	2.00E-57
7647	U13185	Cloning vector pbetagal-Enhancer, complete sequence.	2.00E-79
7648	U55939	Expression vector pVP-Nco, complete sequence.	1.00E-76
7649	D87671	Rattus norvegicus mRNA for TIP120, complete cds	1.00E-87
7650	U25691	Mus musculus lymphocyte specific helicase mRNA, complete cds	4.00E-86
7651	U55939	Expression vector pVP-Nco, complete sequence.	5.00E-79
7652	Z12112	pWE15A cosmid vector DNA	2.00E-79

7653	U13185	Cloning vector pbetagal-Enhancer, complete sequence.	2.00E-79
7654	U13185	Cloning vector pbetagal-Enhancer, complete sequence.	6.00E-80
7655	Z12112	pWE15A cosmid vector DNA	6.00E-80
7656	U09661	Cloning vector pSEAP-Control, complete sequence	6.00E-80
7657	U36909	Bos taurus Rho-associated kinase mRNA, complete cds	2.00E-90
7658	L36610	Mus musculus protein synthesis initiation factor 4A (elF-4A) gene, exons 5, 6, 7, 8, and 9.	2.00E-71
7659	S79463	M-Sema F=a factor in neural network development	1.00E-85
7660	U35312	Mus musculus nuclear receptor co-repressor mRNA, complete cds	1.00E-98
7667	L32977	Homo sapiens (clone f17252) ubiquinol cytochrome c reductase Rieske iron-sulphur protein (UQCRFS1) gene, exon 2	0
7672	S78454	Mus musculus metal response element DNA-binding protein M96 mRNA, complete cds	0
7682	M88458	Human ELP-1 mRNA sequence.	0
7718	S77512	LAMB2=laminin beta 2 chain [human, placenta, mRNA, 5642 nt]	e-131
7720	X53305	H.sapiens mRNA for stathmin	0
7721	J03591	Human ADP/ATP translocase mRNA, 3' end, clone pHAT3.	0
7726	L18964	Human protein kinase C iota isoform (PRKCI) mRNA, complete cds.	2E-67
7736	D29956	Human mRNA for KIAA0055 gene, complete cds	0
7745	M26697	Human nucleolar protein (B23) mRNA, complete cds.	e-149
7765	U47322	Cloning vector DNA, complete sequence.	4E-65
7785	NM_002079.1	Homo sapiens glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1) aspartate aminotransferase mRNA, complete cds.	0

7789	U55939	Expression vector pVP-Nco, complete sequence.	2E-70
7790	D80007	Human mRNA for KIAA0185 gene, partial cds	0
7791	NM_001904.1	Homo sapiens catenin (cadherin-associated protein), beta 1 (88kD) (CTNNB1) mRNA > :: emb|X87838|HSRNABECA H.sapiens mRNA for beta-catenin	e-108
7797	U19867	Cloning vector pSPL3, exon splicing vector, complete sequence, HIV envelope protein gp160 and beta-lactamase, complete cds.	1E-44
7798	M31061	Human ornithine decarboxylase gene, complete cds.	0
7817	Z96177	H.sapiens telomeric DNA sequence, clone 10QTEL040, read 10QTELOO040.seq	2E-70
7818	NM_001904.1	Homo sapiens catenin (cadherin-associated protein), beta 1 (88kD) (CTNNB1) mRNA > :: emb|X87838|HSRNABECA H.sapiens mRNA for beta-catenin	e-176
7854	X83577	M.musculus mRNA for K-glypican	0
7857	S79539	Pat-12=Pat-12 product [mice, embryonic stem ES cells, mRNA, 2781 nt]	e-176
7869	L38951	Homo sapiens importin beta subunit mRNA, complete cds	1E-78
7872	NM_003902.1	Homo sapiens far upstream element binding protein (FUBP) mRNA > :: gb|U05040|HSU05040 Human FUSE binding protein mRNA, complete cds.	0
7887	L08783	BlueScribe M13 Plus cloning vector.	0
7905	U86751	Human nucleolar fibrillar center protein (ASE-1) mRNA, complete cds	8E-95
7913	M21188	Human insulin-degrading enzyme (IDE) mRNA, complete cds.	e-134
7927	NM_001614.1	Homo sapiens actin, gamma 1 (ACTG1) mRNA > :: emb|X04098|HSACTCGR Human mRNA for cytoskeletal gamma-actin	0.00E+00

7932	U12404	Human Csa-19 mRNA, complete cds.	0
7933	X79236	H.sapiens rps26 gene	e-145
7934	NM_003313.1	Homo sapiens tissue specific transplantation antigen P35B (TSTA3) mRNA > :: gb|U58766|HSU58766 Human FX protein mRNA, complete cds	0
7935	M27436	Human tissue factor gene, complete cds, with a Alu repetitive sequence in the 3' untranslated region. > :: gb|I05724| Sequence 12 from Patent EP 0278776	e-121
7945	X79067	H.sapiens ERF-1 mRNA 3' end	0
7946	NM_003017.1	Homo sapiens splicing factor, arginine/serine-rich 3 (SFRS3) mRNA > :: gb|L10838|HUMSRP20 Homo sapiens SR protein family, pre-mRNA splicing factor (SRp20) mRNA, complete cds.	e-135
7953	U48807	Human MAP kinase phosphatase (MKP-2) mRNA, complete cds	0.00E+00
7954	U48807	Human MAP kinase phosphatase (MKP-2) mRNA, complete cds	0.00E+00
7969	U04817	Human protein kinase PITSLRE alpha 2-3 mRNA, complete cds.	8.00E-53
7972	U18297	Human MST1 (MST1) mRNA, complete cds.	0.00E+00
7973	NM_001859.1	Homo sapiens solute carrier family 31 gb|U83460|HSU83460 Human high-affinity copper uptake protein (hCTR1) mRNA, complete cds	0
7985	X70272	single stranded replicative centromeric Saccharomyces cerevisiae /E. coli shuttle vector	3.00E-76
7993	L26050	Human mitochondrial 2,4-dienoyl-CoA reductase mRNA, complete cds.	0.00E+00
7995	X06747	Human hnRNP core protein A1	e-157
7997	M64571	Human microtubule-associated protein 4 mRNA, complete cds.	0.00E+00
8004	X65322.1	Cloning vector pCAT-Basic	9.00E-53

8009	NM_002654.1	Homo sapiens pyruvate kinase, muscle (PKM2) mRNA > :: gb|M23725|HUMPKM2L Human M2-type pyruvate kinase mRNA, complete cds.	e-159
8012	U49352	Human liver 2,4-dienoyl-CoA reductase mRNA, complete cds	2.00E-71
8022	D31889	Human mRNA for KIAA0072 gene, partial cds > :: gb|G27027|G27027 human STS SHGC-31585.	e-167
8037	U43944	Human breast cancer cytosolic NADP(+)-dependent malic enzyme mRNA, partial cds	1.00E-89
8067	U83659	Human multidrug resistance-associated protein homolog (MRP3) mRNA, partial cds	3.00E-85
8092	M33519	Human HLA-B-associated transcript 3 (BAT3) mRNA, complete cds.	3.00E-84
8093	U55387	Cricetulus griseus SL15 mRNA, complete cds	e-150
8114	L36315	Mus musculus (clone pMLZ-1) zinc finger protein	e-162
8121	NM_003902.1	Homo sapiens far upstream element binding protein (FUBP) mRNA > :: gb|U05040|HSU05040 Human FUSE binding protein mRNA, complete cds.	e-175
8128	X56932	H.sapiens mRNA for 23 kD highly basic protein	0.00E+00
8135	X98654	H.sapiens mRNA for DRES9 protein	9.00E-97
8146	S62077	HP1Hs alpha=25 kda chromosomal autoantigen [human, mRNA, 876 nt]	4.00E-68
8153	M23619	Human HMG-I protein isoform mRNA (HMGI gene), clone 6A.	e-117
8173	NM_003217.1	Homo sapiens testis enhanced gene transcript	4E-99
8188	U18671	Human Stat2 gene, complete cds.	0.00E+00
8192	D43636	Human mRNA for KIAA0096 gene, partial cds	0
8194	NM_002734.1	Homo sapiens protein kinase, cAMP-dependent, regulatory, type I, alpha (tissue specific extinguisher 1) (PRKAR1A) mRNA > :: gb|M33336|HUMCAMPPK Human cAMP-dependent protein kinase type I-alpha subunit	0

8195	U72621	Human LOT1 mRNA, complete cds	0.00E+00
8208	NM_003902.1	Homo sapiens far upstream element binding protein (FUBP) mRNA > :: gb|U05040|HSU05040 Human FUSE binding protein mRNA, complete cds.	0.00E+00
8214	L41142	Homo sapiens signal transducer and activator of transcription (STAT5) mRNA, complete cds.	0.00E+00
8215	Z48950	H.sapiens hH3.3B gene for histone H3.3	0.00E+00
8249	L09260	Human (chromosome 3p25) membrane protein mRNA.	e-100
8254	X65304.1	Cloning vector pGEM-3Z	e-173
8259	NM_003358.1	Homo sapiens UDP-glucose ceramide glucosyltransferase (UGCG) mRNA > :: dbj|D50840|HUMCGA Homo sapiens mRNA for ceramide glucosyltransferase, complete cds > :: dbj|E12454|E12454 cDNA encoding human ceramide glucosyltransferase	e-141
8275	M95605	Bos taurus S-adenosylmethionine decarboxylase	e-175
8276	M12623	Human non-histone chromosomal protein HMG-17 mRNA, complete cds.	0.00E+00
8277	U79143	Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA, complete cds	0.00E+00
8288	K01906	Human fetal liver c-myc proto-oncogene, exon 3 and flanks.	e-165
8290	X74870	H.sapiens gene for RNA pol II largest subunit, exons 23-29	e-161
8331	L16991	Human thymidylate kinase (CDC8) mRNA, complete cds.	0.00E+00
8353	L08783	BlueScribe M13 Plus cloning vector.	0.00E+00
8372	NM_002245.1	Homo sapiens potassium inwardly-rectifying channel, subfamily K, member 1 (KCNK1) mRNA > :: gb|U33632|HSU33632 Human two P-domain K+ channel TWIK-1 mRNA, complete cds.	0

8374	D50734	Rat mRNA of antizyme inhibitor, complete cds	e-157
8375	U26401	Human galactokinase (galK) mRNA, complete cds. >	0.00E+00
8381	U49058	Rattus norvegicus CTD-binding SR-like protein rA4 mRNA, partial cds	e-138
8383	X65306.1	Cloning vector pGEM-3Zf(+)	e-116
8395	NM_001172.1	Homo sapiens arginase, type II (ARG2) mRNA > :: gb|U82256|HSU82256 Homo sapiens arginase type II mRNA, complete cds	e-127
8405	M25160	Human Na,K-ATPase beta subunit (ATP1B) gene, exons 3 through 6.	0.00E+00
8411	Y08736	H.sapiens vegf gene, 3'UTR	1.00E-78
8416	U13737	Human cysteine protease CPP32 isoform alpha mRNA, complete cds.	0.00E+00
8419	Y08135	M.musculus mRNA for ASM-like phosphodiesterase 3a	e-148
8420	Y08135	M.musculus mRNA for ASM-like phosphodiesterase 3a	0
8424	NM_001677.1	Homo sapiens ATPase, Na+/K+ transporting, beta 1 polypeptide (ATP1B1) mRNA > :: emb|X03747|HSATPBR Human mRNA for Na/K-ATPase beta subunit	1E-77
8433	Y08135	M.musculus mRNA for ASM-like phosphodiesterase 3a	e-168
8460	U54778	Human 14-3-3 epsilon mRNA, complete cds	1E-67
8461	Y08135	M.musculus mRNA for ASM-like phosphodiesterase 3a	0
8464	NM_001172.1	Homo sapiens arginase, type II (ARG2) mRNA > :: gb|U82256|HSU82256 Homo sapiens arginase type II mRNA, complete cds	e-127
8481	AB002293	Human mRNA for KIAA0295 gene, partial cds	0
8490	M21188	Human insulin-degrading enzyme (IDE) mRNA, complete cds.	2E-81

8521	D87466	Human mRNA for KIAA0276 gene, partial cds	1E-97
8525	U58884	Mus musculus SH3-containing protein SH3P7 mRNA, complete cds. similar to Human Drebrin	4E-96
8537	AB005216	Homo sapiens mRNA for Nck, Ash and phospholipase C gamma-binding protein NAP4, partial cds	0
8538	NM_001960.1	Homo sapiens eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein) (EEF1D) mRNA > :: emb|Z21507|HSEF1DELA H.sapiens EF-1delta gene encoding human elongation factor-1-delta	0.00E+00
8540	M92449	Human LTR mRNA, 3' end of coding region and 3' flank.	e-143
8548	NM_003350.1	Homo sapiens ubiquitin-conjugating enzyme E2 variant 2 (UBE2V2) mRNA > :: emb|X98091|HSVITDITR H.sapiens mRNA for protein induced by vitamin D	0
8552	U44975	Homo sapiens DNA-binding protein CPBP (CPBP) mRNA, partial cds	5.00E-69
8555	Z84510	H.sapiens flow-sorted chromosome 6 HindIII fragment, SC6pA28B7	4.00E-66
8559	Z48042	H.sapiens mRNA encoding GPI-anchored protein p137	e-172
8593	U32986	Human xeroderma pigmentosum group E UV-damaged DNA binding factor mRNA, complete cds.	0
8611	NM_003419.1	Homo sapiens zinc finger protein 10 (KOX 1) for zinc finger protein	e-129
8616	Y00711	Human mRNA for lactate dehydrogenase B (LDH-B)	0.00E+00
8622	Y10725	M.musculus mRNA for protein kinase KIS	0.00E+00
8639	X62078	H.sapiens mRNA for GM2 activator protein	e-164
8644	NM_001009.1	Homo sapiens ribosomal protein S5 (RPS5) mRNA complete cds.	0.00E+00
8652	U97188	Homo sapiens putative RNA binding protein KOC	1E-86

8671	NM_002852.1	Homo sapiens pentaxin-related gene, rapidly induced by IL-1 beta (PTX3) mRNA > :: emb|X63613|HSPTX3R H.sapiens mRNA for pentaxin (PTX3)	0.00E+00
8674	X67155	H.sapiens mRNA for mitotic kinesin-like protein-1	0.00E+00
8684	M54968	Human K-ras oncogene protein mRNA, complete cds >	e-123
8687	D88687	Homo sapiens mRNA for KM-102-derived reductase-like factor, complete cds	0
8689	NM_001436.1	Homo sapiens fibrillarin (FBL) mRNA > :: gb|M59849|HUMFIBAA Human fibrillarin (Hfib1) mRNA, complete cds.	e-103
8691	AB002326	Human mRNA for KIAA0328 gene, partial cds	0.00E+00
8694	M11948	Human promyelocytic leukemia cell mRNA, clones pHH58 and pHH81.	9.00E-84

Table 41B Nearest Neighbor (BlastX vs. Non-Redundant Proteins)

SEQ ID	ACC'N	DESCRIP.	P VALUE
6133	4239895	(AB016816) MASL1 [Homo sapiens]	9.00E-54
6162	4514653	(AB024057) vascular Rab-GAP/TBC-containing protein [Homo sapiens]	6.00E-55
6174	4454524	(AC004841) similar to insulin receptor substrate BAP2; similar to PID:g4126477 [Homo sapiens]	6.00E-22
6175	4545264	(AF118240) peroxisomal biogenesis factor 16 [Homo sapiens]	1.00E-45
6208	3413938	(AB007963) KIAA0494 protein [Homo sapiens]	3.00E-44
6218	4239895	(AB016816) MASL1 [Homo sapiens]	1.00E-47
6235	4502371	breast cancer antiestrogen resistance 3 >gi|3237306 (U92715) breast cancer antiestrogen resistance 3 protein [Homo sapiens]	2.00E-44
6250	4586880	(AB017114) AD 3 [Homo sapiens]	4.00E-48

6253	3327170	(AB014578) KIAA0678 protein [Homo sapiens]	2.00E-51
6264	3153241	(AF053004) class I cytokine receptor [Homo sapiens]	2.00E-17
6267	4138233	(AJ007780) parp-2 gene [Mus musculus]	2.00E-32
6270	3287173	(AJ006266) AND-1 protein [Homo sapiens]	2.00E-42
6283	4507145	UNKNOWN >gi|3873216 (AF065485) sorting nexin 4 [Homo sapiens]	8.00E-46
6303	4153860	(AC005074) similar to U47321 (PID:g1245146) [Homo sapiens]	4.00E-15
6320	3236430	(AF067379) ubiquitin-protein ligase E3-alpha [Mus musculus]	3.00E-35
6349	3043696	(AB011158) KIAA0586 protein [Homo sapiens]	1.00E-44
6356	4519623	(AB017616) homologous to the yeast YGR163 gene [Mus musculus]	2.00E-54
6376	4455035	(AF116238) pseudouridine synthase 1 [Homo sapiens]	4.00E-48
6400	3075377	(AC004602) F23487_2 [Homo sapiens]	2.00E-21
6402	4505611	poly(A)-specific ribonuclease	7.00E-41
6469	1825606	(U88169) similar to molybdoterin biosynthesis MOEB proteins [Caenorhabditis elegans]	2.00E-37
6478	4586287	(AB004794) DUF140 [Xenopus laevis]	7.00E-45
6492	3941342	(AF043250) mitochondrial outer membrane protein [Homo sapiens] >gi|3941347 (AF043253) mitochondrial outer membrane protein [Homo sapiens] >gi|4105703|gb|AAD02504|	5.00E-40
6510	4586844	(AB015633) type II membrane protein	2.00E-46
6518	3327078	(AB014532) KIAA0632 protein [Homo sapiens]	6.00E-36
6529	3327230	(AB014608) KIAA0708 protein [Homo sapiens]	5.00E-52
6568	3372677	(AF061749) tumorous imaginal discs protein Tid56 homolog	7.00E-35
6598	4050034	(AF098482) transcriptional coactivator p52 [Homo sapiens]	1.00E-36
6600	4406632	(AF131801) Unknown [Homo sapiens]	3.00E-21
6608	3114828	(AJ005897) JM5 [Homo sapiens]	3.00E-44

6626	3766209	(AF071777) IRE1 [Mus musculus]	2.00E-29
6657	3043644	(AB011132) KIAA0560 protein [Homo sapiens]	3.00E-43
6668	3088575	(AF059531) protein arginine N-methyltransferase 3 [Homo sapiens]	4.00E-46
6674	4505891	UNKNOWN >gi|3153235 (AF046889) lysyl hydroxylase isoform 3 [Homo sapiens] >gi|3551836	3.00E-30
6686	3114828	(AJ005897) JM5 [Homo sapiens]	1.00E-24
6688	3242214	(AJ006778) DRIM protein [Homo sapiens]	2.00E-36
6694	4200236	(AL035308) hypothetical protein [Homo sapiens]	8.00E-09
6696	3413892	(AB007934) KIAA0465 protein [Homo sapiens]	2.00E-51
6731	3043626	(AB011123) KIAA0551 protein [Homo sapiens]	3.00E-31
6739	2498864	RRP5 PROTEIN HOMOLOG (KIAA0185) hypothetical protein YM9959.11C of S.cerevisiae. [Homo sapiens]	3.00E-13
6766	3402197	(AJ010014) M96A protein [Homo sapiens]	1.00E-21
6773	2217964	(Z50798) p52 [Gallus gallus]	7.00E-14
6782	3043626	(AB011123) KIAA0551 protein [Homo sapiens]	1.00E-40
6793	135470	TUBULIN BETA-5 CHAIN sapiens]	3.00E-21
6797	3327056	(AB014521) KIAA0621 protein [Homo sapiens]	2.00E-29
6800	4506787	UNKNOWN GTPASE-ACTIVATING-LIKE PROTEIN IQGAP1 (P195) (KIAA0051) protein - human >gi|473931|dbj|BAA06123| (D29640) KIAA0051 [Homo sapiens] >gi|536844 (L33075) ras GTPase-activating-like protein [Homo sapiens]	4.00E-41
6805	1350762	60S RIBOSOMAL PROTEIN L6 sapiens]	2.00E-22
6809	2687400	(AF035824) vesicle soluble NSF attachment protein receptor [Homo sapiens]	1.00E-23
6826	2914385	Chain C, Human Pcna >gi|2914387|pdb|1AXC|E Chain E, Human Pcna	2.00E-27
6827	284076	ERD-2-like protein, ELP-1 - human	1.00E-26

6829	2497524	KINESIN-LIKE PROTEIN KIF1B mouse >gi|407339|dbj|BAA04503| (D17577) Kif1b [Mus musculus]	9.00E-33
6831	3327056	(AB014521) KIAA0621 protein [Homo sapiens]	1.00E-13
6832	279567	insulinase (EC 3.4.99.45) - human	2.00E-26
6834	487416	(L20302) actin filament protein [Gallus gallus]	3.00E-45
6835	1731428	ZINC FINGER PROTEIN ZFP-38	7.00E-35
6836	968973	(U29156) involved in signaling by the epidermal growth factor receptor; Method: conceptual translation supplied by author. [Mus musculus]	1.00E-22
6837	1552350	(Y08135) acid sphingomyelinase-like phosphodiesterase [Mus musculus]	2.00E-35
6838	3327098	(AB014542) KIAA0642 protein [Homo sapiens]	3.00E-15
6839	3914801	DNA-DIRECTED RNA POLYMERASE I 135 KD POLYPEPTIDE (RNA POLYMERASE I SUBUNIT 2) (RPA135) (RNA POLYMERASE I 127 KD SUBUNIT) >gi|2739048 (AF025424) RNA polymerase I 127 kDa subunit [Rattus norvegicus]	2.00E-45
6841	4165018	(D89053) Acyl-CoA synthetase 3 [Homo sapiens]	2.00E-53
6842	968973	(U29156) involved in signaling by the epidermal growth factor receptor; Method: conceptual translation supplied by author. [Mus musculus]	3.00E-40
6843	4519883	(AB017970) dipeptidyl peptidase III	4.00E-50
6844	3327052	(AB014519) KIAA0619 protein [Homo sapiens]	7.00E-30
6845	538413	(L36315) zinc finger protein [Mus musculus]	6.00E-55
6846	1717793	PROTEIN TSG24 (MEIOTIC CHECK POINT REGULATOR) >gi|1083553|pir||A55117 tsg24 protein - mouse	1.00E-50
6847	3420277	(AF064826) glypican 4 [Homo sapiens]	3.00E-54
6904	4580645	(AF118855) trans-prenyltransferase [Mus musculus]	2.00E-48
6925	3882171	(AB018268) KIAA0725 protein [Homo sapiens]	3.00E-24

6929	4104976	(AF043117) ubiquitin-fusion degradation protein 2 [Homo sapiens]	2.00E-41
6937	3242214	(AJ006778) DRIM protein [Homo sapiens]	4.00E-34
7010	4191810	(AB006532) DNA helicase [Homo sapiens]	5.00E-41
7055	3043714	(AB011167) KIAA0595 protein [Homo sapiens]	5.00E-20
7078	4379097	(Y17999) Dyrk1B protein kinase [Homo sapiens]	3.00E-21
7124	3043712	(AB011166) KIAA0594 protein [Homo sapiens]	2.00E-49
7175	4240227	(AB020676) KIAA0869 protein [Homo sapiens]	4.00E-35
7187	4235226	(AF061025) leucine zipper-EF-hand containing transmembrane protein 1 [Homo sapiens]	6.00E-34
7230	3426268	(AF044201) neural membrane protein 35; NMP35 [Rattus norvegicus]	1.00E-29
7248	4507367	threonyl-tRNA synthetase SYNTHETASE, CYTOPLASMIC (THREONINE--TRNA LIGASE) (THRRS) 6.1.1.3) - human >gi|1464742 (M63180) threonyl-tRNA synthetase [Homo sapiens]	3.00E-26
7249	2072294	(U95097) mitotic phosphoprotein 43 [Xenopus laevis]	1.00E-19
7259	543222	glutamine (Q)-rich factor 1, QRF-1 - mouse factor 1, QRF-1 [mice, B-cell leukemia, BCL1, Peptide Partial, 84 aa]	1.00E-39
7260	3335569	(AF072759) fatty acid transport protein 4; FATP4 [Mus musculus]	7.00E-39
7264	2996194	(AF053232) SIK similar protein [Mus musculus]	1.00E-31
7268	2935597	(AC004262) R29368_2 [Homo sapiens]	6.00E-49
7297	2645205	(U63648) p160 myb-binding protein [Mus musculus]	1.00E-21
7300	1407655	(U58884) SH3P7 [Mus musculus]	8.00E-21
7310	2134381	polybromo 1 protein - chicken	8.00E-29
7315	4505613	PRKC, apoptosis, WT1, regulator par-4 [Homo sapiens]	6.00E-34
7325	3757892	(AF079765) enhancer of polycomb [Mus musculus]	3.00E-41
7327	2134436	zinc finger protein - chicken (fragment)	4.00E-37

7328	2393722	(U90313) glutathione-S-transferase homolog [Homo sapiens]	6.00E-34
7330	459002	(U00036) R151.6 gene product [Caenorhabditis elegans]	7.00E-10
7332	119530	PROTEIN DISULFIDE ISOMERASE-RELATED PROTEIN PRECURSOR (ERP72) >gi|87320|pir||A23723 protein disulfide-isomerase (EC 5.3.4.1) ERp72 precursor - human protein [Homo sapiens]	3.00E-23
7335	2073541	(L19437) transaldolase [Homo sapiens] >gi|2612879	2.00E-24
7337	984125	(X90857) -14 [Homo sapiens]	2.00E-23
7341	4106818	(AF083395) phospholipase A2-activating protein [Homo sapiens]	4.00E-36
7343	4507955	YY1 transcription factor REPRESSOR PROTEIN YY1 (YIN AND YANG 1) (YY-1) (DELTA TRANSCRIPTION FACTOR) (NF-E1) >gi|38011|emb|CAA78455|	4.00E-27
7346	1698779	(U70372) PAM COOH-terminal interactor protein 2 [Rattus norvegicus]	6.00E-35
7348	4204684	(AF102542) beta-1,6-N-acetylglucosaminyltransferase core 2/core 4 beta-1,6-N-acetylglucosaminyltransferase; core 2/4-GnT [Homo sapiens]	9.00E-43
7351	2239126	(Y10746) methyl-CpG binding protein [Homo sapiens]	4.00E-16
7355	1747519	(U76759) nuclear protein NIP45 [Mus musculus]	2.00E-29
7356	545790	DARPP-32=dopamine and cAMP-regulated phosphoprotein [human, brain, Peptide, 204 aa] sapiens]	1.00E-29
7357	1709689	PEPTIDE METHIONINE SULFOXIDE REDUCTASE (PEPTIDE MET(O) REDUCTASE) >gi|1205993 taurus]	1.00E-37

7361	2736151	(AF021935) mytonic dystrophy kinase-related Cdc42-binding kinase [Rattus norvegicus]	1.00E-41
7363	3329392	(AF038961) SL15 protein [Homo sapiens]	8.00E-36
7364	4097712	(U67322) HBV associated factor [Homo sapiens]	7.00E-56
7365	585084	ELONGATION FACTOR G, MITOCHONDRIAL PRECURSOR (MEF-G) >gi|543383|pir||S40780 translation elongation factor G, mitochondrial - rat >gi|310102	7.00E-49
7366	1438534	(U49057) rA9 [Rattus norvegicus]	3.00E-45
7367	1336628	(U57368) EGF repeat transmembrane protein [Mus musculus]	7.00E-47
7368	3914802	DNA-DIRECTED RNA POLYMERASE I LARGEST SUBUNIT (RNA POLYMERASE I 194 KD SUBUNIT) (RPA194)	1.00E-37
7369	3387977	(AF070598) ABC transporter [Homo sapiens]	5.00E-50
7370	1717793	PROTEIN TSG24 (MEIOTIC CHECK POINT REGULATOR) >gi|1083553|pir||A55117 tsg24 protein - mouse	2.00E-48
7371	2493735	SKD3 PROTEIN SKD3 [Mus musculus]	7.00E-43
7372	1041038	(D78020) NFI-A4 [Rattus norvegicus]	3.00E-26
7380	4455118	(AF125158) zinc finger DNA binding protein 99	9.00E-41
7418	4049922	(AF072810) transcription factor WSTF [Homo sapiens]	4.00E-48
7434	4586287	(AB004794) DUF140 [Xenopus laevis]	3.00E-45
7441	3435244	(AF083322) centriole associated protein CEP110 [Homo sapiens]	2.00E-40
7466	3413886	(AB007931) KIAA0462 protein [Homo sapiens]	2.00E-35
7558	3882311	(AB018338) KIAA0795 protein [Homo sapiens]	4.00E-47
7593	4240167	(AB020646) KIAA0839 protein [Homo sapiens]	2.00E-46
7613	4191610	(AF117107) IGF-II mRNA-binding protein 2 [Homo sapiens]	3.00E-49
7615	3135669	(AF064084) prenylcysteine carboxyl methyltransferase	1.00E-39

7625	3043548	(AB011084) KIAA0512 protein [Homo sapiens]	2.00E-47
7627	3093476	(AF008915) EVI-5 homolog [Homo sapiens]	6.00E-19
7628	3834629	(AF094519) diaphanous-related formin; p134 mDia2 [Mus musculus]	1.00E-32
7629	3193226	(AF068706) gamma2-adaptin [Homo sapiens]	1.00E-46
7630	3851584	(AF092563) chromosome-associated protein-E [Homo sapiens]	4.00E-48
7631	4101695	(AF006010) progestin induced protein [Homo sapiens]	5.00E-30
7646	3850704	(AJ005273) Kin17 [Homo sapiens]	9.00E-24
7649	4240147	(AB020636) KIAA0829 protein [Homo sapiens]	9.00E-41
7650	2137490	lymphocyte specific helicase - mouse musculus]	5.00E-35
7657	3327052	(AB014519) KIAA0619 protein [Homo sapiens]	1.00E-41
7659	2137494	M-sema F protein precusor - mouse F [mice, neonatal brain, Peptide, 834 aa] [Mus sp.]	7.00E-34
7660	2137603	nuclear receptor co-repressor N-CoR - mouse musculus] >gi|1583865|prf||2121436A thyroid hormone receptor co-repressor [Mus musculus]	9.00E-41
7661	2674107	(AF023451) guanine nucleotide-exchange protein [Bos taurus]	3.00E-48
7683	3659505	(AC005084) similar to mouse mCASK-A; similar to e1288039	1.00E-57
7745	114762	NUCLEOPHOSMIN (NPM) (NUCLEOLAR PHOSPHOPROTEIN B23) (NUMATRIN) (NUCLEOLAR PROTEIN NO38) sapiens]	6.00E-35
7747	3327056	(AB014521) KIAA0621 protein [Homo sapiens]	8.00E-40
7784	4545264	(AF118240) peroxisomal biogenesis factor 16 [Homo sapiens]	2.00E-65
7790	2498864	RRP5 PROTEIN HOMOLOG (KIAA0185) hypothetical protein YM9959.11C of S.cerevisiae. [Homo sapiens]	7.00E-77
7854	3420277	(AF064826) glypican 4 [Homo sapiens]	4.00E-76

7864	3088575	(AF059531) protein arginine N-methyltransferase 3 [Homo sapiens]	7.00E-97
7867	4050034	(AF098482) transcriptional coactivator p52 [Homo sapiens]	2.00E-58
7907	4506357	UNKNOWN; PZR >gi|3851145 sapiens]	2.00E-60
7926	3387977	(AF070598) ABC transporter [Homo sapiens]	e-113
7932	1709974	60S RIBOSOMAL PROTEIN L10A protein L10a [Rattus norvegicus] Ribosomal Protein RPL10A) [Homo sapiens]	e-111
7934	4507709	tissue specific transplantation antigen P35B >gi|1381179 (U58766) FX [Homo sapiens]	9.00E-90
7972	1117791	(U18297) MST1 [Homo sapiens]	4E-85
7973	4507015	copper transporter 1	3.00E-72
7993	4503301	2,4-dienoyl CoA reductase REDUCTASE, MITOCHONDRIAL PRECURSOR (2,4-DIENOYL-COA REDUCTASE (NADPH)) (4-ENOYL-COA REDUCTASE (NADPH)) precursor, mitochondrial - human >gi|602703 (L26050) 2,4-dienoyl-CoA reductase [Homo sapiens] >gi|2673979 precursor [Homo sapiens] >gi|4126313 (AF049895) 2,4-dienoyl-CoA reductase [Homo sapiens]	6E-94
7997	126743	MICROTUBULE-ASSOCIATED PROTEIN 4 human >gi|187383 (M64571) microtubule-associated protein 4 [Homo sapiens]	6E-84
8010	4505987	PTPRF interacting protein, binding protein 1 (liprin beta 1) >gi|3309539 (AF034802) liprin-beta1 [Homo sapiens]	4E-89
8016	3043644	(AB011132) KIAA0560 protein [Homo sapiens]	e-108
8040	3413892	(AB007934) KIAA0465 protein [Homo sapiens]	7.00E-87
8052	4185796	(AF103796) placenta-specific ATP-binding cassette transporter [Homo sapiens]	2E-68

8069	4507145	UNKNOWN >gi|3873216 (AF065485) sorting nexin 4 [Homo sapiens]	1.00E-73
8104	1083566	zinc finger protein/transactivator Zfp-38 - mouse >gi|55477 |emb| CAA45280| (X63747) Zfp-38 [Mus musculus]	2E-64
8114	1806134	(Z67747) zinc finger protein [Mus musculus]	7.00E-78
8128	730451	60S RIBOSOMAL PROTEIN L13A (23 KD HIGHLY BASIC PROTEIN) >gi|345897|pir||S29539 basic protein, 23K - human >gi|23691|emb|CAA40254| (X56932) 23 kD highly basic protein [Homo sapiens]	4.00E-87
8381	4102967	(AF023142) pre-mRNA splicing SR protein rA4 [Homo sapiens]	1.00E-33
8413	3108093	(AF061258) LIM protein [Homo sapiens]	6.00E-82
8414	3170887	(AF061555) ubiquitin-protein ligase E3-alpha [Mus musculus]	e-104
8420	1552350	(Y08135) acid sphingomyelinase-like phosphodiesterase [Mus musculus]	6.00E-91
8461	1552350	(Y08135) acid sphingomyelinase-like phosphodiesterase [Mus musculus]	e-106
8462	3242214	(AJ006778) DRIM protein [Homo sapiens]	e-114
8483	4514653	(AB024057) vascular Rab-GAP/TBC-containing protein [Homo sapiens]	e-121
8537	2443367	(AB005216) Nck, Ash and phospholipase C gamma-binding protein NAP4 [Homo sapiens]	e-120
8571	119110	EBNA-1 NUCLEAR PROTEIN herpesvirus 4 (strain B95-8) >gi|1334880|emb|CAA24816.1| gene. [Human herpesvirus 4]	2.00E-38
8575	121640	GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN PRECURSOR >gi|72320|pir||KNMU glycine-rich cell wall protein precursor - Arabidopsis thaliana	8.00E-31
8591	1362077	glycin-rich protein - cowpea (fragment)	2E-40

8615	121640	GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN PRECURSOR >gi|72320|pir||KNMU glycine-rich cell wall protein precursor - Arabidopsis thaliana	9.00E-27
8642	2674107	(AF023451) guanine nucleotide-exchange protein [Bos taurus]	5E-89
8644	3717978	(Y12431) 5S ribosomal protein [Mus musculus]	5E-94
8652	4191610	(AF117107) IGF-II mRNA-binding protein 2 [Homo sapiens]	e-111
8674	2119281	CHO1 antigen - Chinese hamster	e-101
8675	3435244	(AF083322) centriole associated protein CEP110 [Homo sapiens]	2E-70
8687	1843434	(D88687) KM-102-derived reductase-like factor [Homo sapiens]	4.00E-91
8700	3834629	(AF094519) diaphanous-related formin; p134 mDia2 [Mus musculus]	1E-49

Example 29: Members of Protein Families

SEQ ID NOS: 7662-8706 were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein family (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 42A, inserted prior to claims). Table 42A provides the SEQ ID NO: of the query sequence, a brief description of the profile hit, the position of the query sequence within the individual sequence (indicated as "start" and "stop"), and the orientation (Direction) of the query sequence with respect to the individual sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Table 42A Profile Hits

SEQ ID NO:	Description	Start	Stop	Dir
8063	14_3_3 proteins	166	845	for
8462	3'5'-cyclic nucleotide phosphodiesterases	64	573	for
7675	4 transmembrane integral membrane proteins	300	924	rev
8074	4 transmembrane integral membrane proteins	340	941	rev
7748	7 transmembrane receptor (rhodopsin family)	109	647	rev
8023	7 transmembrane receptor (rhodopsin family)	84	947	rev
8164	7 transmembrane receptor (rhodopsin family)	305	975	for
7694	7 transmembrane receptor (Secretin family)	50	1269	for
7815	7 transmembrane receptor (Secretin family)	63	1160	rev
8007	7 transmembrane receptor (Secretin family)	38	869	rev
8023	7 transmembrane receptor (Secretin family)	237	930	rev
8164	7 transmembrane receptor (Secretin family)	188	975	for
8437	7 transmembrane receptor (Secretin family)	377	1524	rev
7767	ATPases Associated with Various Cellular Activities	136	718	for
7768	ATPases Associated with Various Cellular Activities	271	765	for

7784	ATPases Associated with Various Cellular Activities	206	709	rev
7892	ATPases Associated with Various Cellular Activities	139	783	for
7926	ATPases Associated with Various Cellular Activities	265	713	for
7968	ATPases Associated with Various Cellular Activities	152	616	rev
8009	ATPases Associated with Various Cellular Activities	12	510	for
8018	ATPases Associated with Various Cellular Activities	125	658	for
8060	ATPases Associated with Various Cellular Activities	97	752	for
8093	ATPases Associated with Various Cellular Activities	185	664	for
8128	ATPases Associated with Various Cellular Activities	69	485	for
8266	ATPases Associated with Various Cellular Activities	73	550	for
8273	ATPases Associated with Various Cellular Activities	340	928	for
8386	ATPases Associated with Various Cellular Activities	872	1390	rev
8439	ATPases Associated with Various Cellular Activities	122	635	for
8454	ATPases Associated with Various Cellular Activities	84	492	rev
8486	ATPases Associated with Various Cellular Activities	31	434	rev
8510	ATPases Associated with Various Cellular Activities	953	1358	rev

8557	ATPases Associated with Various Cellular Activities	192	690	rev
8572	ATPases Associated with Various Cellular Activities	51	593	for
8578	ATPases Associated with Various Cellular Activities	135	615	rev
8674	ATPases Associated with Various Cellular Activities	0	673	for
7719	Basic region plus leucine zipper transcription factors	81	277	for
7811	C2 domain (prot. kinase C like)	403	582	for
8522	C2 domain (prot. kinase C like)	493	637	for
8334	Cysteine proteases	359	984	rev
7726	DEAD and DEAH box helicases	34	690	rev
7961	DEAD and DEAH box helicases	43	753	for
8613	DEAD and DEAH box helicases	426	719	for
7810	Dual specificity phosphatase, catalytic domain	365	696	rev
7824	Dual specificity phosphatase, catalytic domain	243	597	for
8183	Dual specificity phosphatase, catalytic domain	786	1566	for
7691	EF-hand	556	630	for
7767	Eukaryotic aspartyl proteases	116	763	for
7874	Eukaryotic aspartyl proteases	92	1008	rev
7999	Eukaryotic aspartyl proteases	73	603	rev
8041	Eukaryotic aspartyl proteases	147	694	rev
8059	Eukaryotic aspartyl proteases	38	740	rev
8087	Eukaryotic aspartyl proteases	404	1113	rev
8226	Eukaryotic aspartyl proteases	237	829	rev
8234	Eukaryotic aspartyl proteases	117	729	rev
8289	Eukaryotic aspartyl proteases	217	1397	rev

8386	Eukaryotic aspartyl proteases	413	1366	rev
8387	Eukaryotic aspartyl proteases	8	710	rev
8444	Eukaryotic aspartyl proteases	291	1146	rev
8526	Eukaryotic aspartyl proteases	216	1158	rev
8592	Eukaryotic aspartyl proteases	228	659	for
8619	Eukaryotic aspartyl proteases	276	1291	rev
8685	Eukaryotic aspartyl proteases	525	1431	for
8064	Fibronectin type II domain	455	565	rev
7875	G-protein alpha subunit	24	583	rev
7717	Helicases conserved C-terminal domain	160	309	for
7748	Helicases conserved C-terminal domain	363	560	rev
8288	Helix-loop-helix DNA binding domain	224	382	for
8277	kinase domain of tors	474	713	for
7921	mkk like kinases	17	626	rev
7972	mkk like kinases	35	719	for
8135	mkk like kinases	114	527	for
8622	mkk like kinases	9	463	for
7878	Neurotransmitter-gated ion-channel	267	1411	for
8018	Neurotransmitter-gated ion-channel	367	1168	for
8164	Neurotransmitter-gated ion-channel	222	1024	for
8198	Neurotransmitter-gated ion-channel	352	1273	for
8250	Neurotransmitter-gated ion-channel	377	1159	for
8634	Neurotransmitter-gated ion-channel	112	1120	for
7717	protein kinase	153	743	for
7726	protein kinase	123	904	for
7801	protein kinase	471	1072	for
7802	protein kinase	190	609	for
7806	protein kinase	235	641	for
7840	protein kinase	8	711	rev
7863	protein kinase	90	537	for
7872	protein kinase	200	524	rev

7878	protein kinase	706	1331	for
7918	protein kinase	24	666	for
7921	protein kinase	56	593	rev
7940	protein kinase	263	824	for
7946	protein kinase	217	779	for
7972	protein kinase	290	711	for
8073	protein kinase	38	776	for
8147	protein kinase	14	657	for
8208	protein kinase	202	644	rev
8265	protein kinase	1	656	for
8301	protein kinase	57	689	for
8338	protein kinase	33	646	for
8387	protein kinase	630	1148	rev
8550	protein kinase	49	761	rev
8622	protein kinase	0	463	for
8654	protein kinase	77	590	for
7815	Protein Tyrosine Phosphatase	82	482	rev
7865	Protein Tyrosine Phosphatase	71	461	rev
8158	Protein Tyrosine Phosphatase	270	704	for
8293	Protein Tyrosine Phosphatase	359	851	for
8371	Protein Tyrosine Phosphatase	56	680	for
7946	RNA recognition motif. (aka RRM, RBD, or RNP domain)	165	365	for
8290	RNA recognition motif. (aka RRM, RBD, or RNP domain)	37	174	for
8537	SH2 Domain	201	362	for
7714	Thioredoxins	253	554	for
7675	Trypsin	252	1007	rev
8386	Trypsin	350	1164	rev
8437	Trypsin	447	1211	rev
8517	Trypsin	14	765	rev

8526	Trypsin	700	1556	rev
8534	Trypsin	47	670	rev
8377	WD domain, G-beta repeats	70	161	for
7675	wnt family of developmental signaling proteins	282	1017	rev
7749	wnt family of developmental signaling proteins	154	978	rev
7874	wnt family of developmental signaling proteins	38	858	rev
7922	wnt family of developmental signaling proteins	574	1318	rev
7971	wnt family of developmental signaling proteins	578	1313	rev
8000	wnt family of developmental signaling proteins	205	1068	rev
8088	wnt family of developmental signaling proteins	2	824	rev
8100	wnt family of developmental signaling proteins	621	1420	rev
8225	wnt family of developmental signaling proteins	394	1343	rev
8241	wnt family of developmental signaling proteins	162	1027	rev
8300	wnt family of developmental signaling proteins	274	1405	rev
8334	wnt family of developmental signaling proteins	560	1195	rev
8386	wnt family of developmental signaling proteins	250	1273	rev
8387	wnt family of developmental signaling proteins	523	1409	rev

8390	wnt family of developmental signaling proteins	297	1237	rev
8437	wnt family of developmental signaling proteins	51	1002	rev
8439	wnt family of developmental signaling proteins	28	1180	rev
8444	wnt family of developmental signaling proteins	638	1614	rev
8469	wnt family of developmental signaling proteins	30	1078	rev
8505	wnt family of developmental signaling proteins	4	1074	rev
8506	wnt family of developmental signaling proteins	208	1107	rev
8510	wnt family of developmental signaling proteins	242	1068	rev
8517	wnt family of developmental signaling proteins	159	1057	rev
8526	wnt family of developmental signaling proteins	844	1691	rev
8532	wnt family of developmental signaling proteins	107	784	rev
8534	wnt family of developmental signaling proteins	127	1226	rev
8559	wnt family of developmental signaling proteins	5	704	rev
8569	wnt family of developmental signaling proteins	328	1193	rev
8607	wnt family of developmental signaling proteins	341	1222	rev
8619	wnt family of developmental signaling proteins	820	1617	rev

8624	wnt family of developmental signaling proteins	461	1283	rev
7831	Zinc finger, C2H2 type	495	557	for
8038	Zinc finger, C2H2 type	500	562	for
8114	Zinc finger, C2H2 type	279	341	for
8350	Zinc finger, C2H2 type	148	210	for
8611	Zinc finger, C2H2 type	422	484	for

Table 42B Profile Hits for Contigs				
SEQ ID NO:	Description	Start	Stop	Dir
8737	ATPases Associated with Various Cellular Activities	118	661	for
8751	ATPases Associated with Various Cellular Activities	135	536	for
8781	ATPases Associated with Various Cellular Activities	142	574	for
8744	DEAD and DEAH box helicases	66	931	rev
8782	Helicases conserved C-terminal domain	51	242	for
8757	Neurotransmitter-gated ion-channel	169	738	rev
8736	Protein phosphatase 2A regulatory subunit PR55	275	1510	for
8751	Protein phosphatase 2A regulatory subunit PR55	55	1087	for
8766	Protein phosphatase 2A regulatory subunit PR55	13	1183	for
8780	Protein phosphatase 2A regulatory subunit PR55	511	1861	rev
8775	Protein Tyrosine Phosphatase	292	768	for
8764	Thioredoxins	182	475	for

Some polynucleotides exhibited multiple profile hits where the query sequence contains overlapping profile regions, and/or where the sequence contains two different functional domains. Each of the profile hits of Table 42A are described in more detail below. The acronyms for the profiles (provided in parentheses) are those used to identify the profile in the Pfam and Prosite databases. The Pfam database can be accessed through many URLS. The Prosite database can be accessed at the Expasy website. The public information available on the Pfam and Prosite databases regarding the various profiles, including but not limited to the activities, function, and consensus sequences of various proteins families and protein domains, is incorporated herein by reference.

14-3-3 Family (14_3_3). Some SEQ ID NOS corresponds to a sequence encoding a 14-3-3 protein family member. The 14-3-3 protein family includes a group of closely related acidic homodimeric proteins of about 30 kD first identified as very abundant in mammalian brain tissues and located preferentially in neurons (Aitken et al. *Trends Biochem. Sci.* (1995) *20*:95-97; Morrison *Science* (1994) *266*:56-57; and Xiao et al. *Nature* (1995) *376*:188-191). The 14-3-3 proteins have multiple biological activities, including a key role in signal transduction pathways and the cell cycle. 14-3-3 proteins interact with kinases (*e.g.*, PKC or Raf-1), and can also function as protein-kinase dependent activators of tyrosine and tryptophan hydroxylases. The 14-3-3 protein sequences are extremely well conserved, and include two highly conserved regions: the first is a peptide of 11 residues located in the N-terminal section; the second, a 20 amino acid region located in the C-terminal section.

3'5'-Cyclin Nucleotide Phosphodiesterases (PDEase). Some SEQ ID NOS represent a polynucleotide encoding a novel 3'5'-cyclic nucleotide phosphodiesterase. PDEases catalyze the hydrolysis of cAMP or cGMP to the corresponding nucleoside 5' monophosphates (Charbonneau et al, *Proc. Natl. Acad. Sci. U.S.A.* (1986) *83*:9308). There are at least seven different subfamilies of PDEases (Beavo et al., *Trends Pharmacol. Sci.* (1990) *11*:150; http://weber.u.washington.edu/~pde/: 1) Type 1, calmodulin/calcium-dependent PDEases; 2) Type 2, cGMP-stimulated PDEases; 3) Type 3, cGMP-inhibited PDEases; 4) Type 4, cAMP-specific PDEases.; 5) Type 5, cGMP-specific PDEases; 6) Type 6, rhodopsin-sensitive cGMP-specific PDEases; and 7) Type 7, High affinity cAMP-specific PDEases. All PDEase forms share a conserved domain of about 270 residues.

Four Transmembrane Integral Membrane Proteins (transmembrane4). Some SEQ

ID NOS correspond to a sequence encoding a member of the four transmembrane segments integral membrane protein family (tm4 family). The tm4 family of proteins includes a number of evolutionarily-related eukaryotic cell surface antigens (Levy *et al.*, *J. Biol. Chem.*, (1991) *266*:14597; Tomlinson *et al.*, *Eur. J. Immunol.* (1993) *23*:136; Barclay *et al.* The leucocyte antigen factbooks. (1993) Academic Press, London/San Diego). The tm4 family members are type III membrane proteins, which are integral membrane proteins containing an N-terminal membrane-anchoring domain that functions both as a translocation signal and as a membrane anchor. The family members also contain three additional transmembrane regions, at least seven conserved cysteines residues, and are of approximately the same size (218 to 284 residues). The consensus pattern spans a conserved region including two cysteines located in a short cytoplasmic loop between two transmembrane domains:

<u>Seven Transmembrane Integral Membrane Proteins -- Rhodopsin Family (7tm_1).</u> Some SEQ ID NOS correspond to a sequence encoding a member of the seven transmembrane (7tm) receptor rhodopsin family. G-protein coupled receptors of the (7tm) rhodopsin family include hormones, neurotransmitters, and light receptors that transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins (Strosberg *Eur. J. Biochem.* (1991) *196*:1, Kerlavage *Curr. Opin. Struct. Biol.* (1991) *1*:394, Probst, et al., *DNA Cell Biol.* (1992) *11*:1, Savarese, et al., Biochem. J. (1992) *283*:1)

<u>Seven Transmembrane Integral Membrane Proteins -- Secretin Family (7tm_2).</u> Some SEQ ID NOS correspond to a sequence encoding a member of the seven transmembrane receptor (7tm) secretin family (Jueppner et al. *Science* (1991) *254*:1024; Hamann et al. *Genomics* (1996) *32*:144). The N-terminal extracellular domain of these receptors contains five conserved cysteines residues involved in disulfide bonds, with a consensus pattern in the region that spans the first three cysteines. One of the most highly conserved regions spans the C-terminal part of the last transmembrane region and the beginning of the adjacent intracellular region and is used as a second signature pattern.

<u>ATPases Associated with Various Cellular Activities (ATPases).</u> Several of the polynucleotides of the invention correspond to a sequence that encodes a member of a family of ATPases Associated with diverse cellular Activities (AAA). The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids containing an ATP-binding site (Froehlich *et al., J. Cell Biol.* (1991) *114*:443; Erdmann *et al. Cell* (1991) *64*:499; Peters *et al., EMBO J.* (1990) *9*:1757; Kunau

et al., *Biochimie* (1993) *75*:209-224; Confalonieri *et al.*, *BioEssays* (1995) *17*:639). The AAA domain, which can be present in one or two copies, acts as an ATP-dependent protein clamp (Confalonieri *et al.* (1995) *BioEssays 17*:639) and contains a highly conserved region located in the central part of the domain.

Basic Region Plus Leucine Zipper Transcription Factors (BZIP). One SEQ ID NOrepresents a polynucleotide encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization.

C2 domain (C2). Some SEQ ID NOS correspond to a sequence encoding a C2 domain, which is involved in calcium-dependent phospholipid binding (Davletov *J. Biol. Chem.* (1993) *268*:26386-26390) or, in proteins that do not bind calcium, the domain may facilitate binding to inositol-1,3,4,5-tetraphosphate (Fukuda et al. *J. Biol. Chem.* (1994) *269*:29206-29211; Sutton et al. *Cell* (1995) *80*:929-938).

Cysteine proteases (Cys-protease). One SEQ ID NO represents a polynucleotide encoding a protein having a eukaryotic thiol (cysteine) protease active site. Cysteine proteases (Dufour *Biochimie* (1988) *70*:1335) are a family of proteolytic enzymes that contain an active site cysteine. Catalysis proceeds through a thioester intermediate and is facilitated by a nearby histidine side chain; an asparagine completes the essential catalytic triad.

DEAD and DEAH box families ATP-dependent helicases (Dead_box_helic). Some SEQ ID NOS represent polynucleotides encoding a novel member of the DEAD and DEAH box families (Schmid et al., *Mol. Microbiol.* (1992) *6*:283; Linder et al., *Nature* (1989) *337*:121; Wassarman, et al., *Nature* (1991) *349*:463). All members of these families are involved in ATP-dependent, nucleic-acid unwinding. All DEAD box family members share a number of conserved sequence motifs, some of which are specific to the DEAD family, with others shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman *Nature* (1988) *333*:22 and *Nature* (1988) *333*:578 (Errata); http://www.expasy.ch/ www/ linder/ HELICASES_ TEXT.html). One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Proteins that have His instead of the second Asp and are 'D-E-A-H-box' proteins (Wassarman et al., *Nature* (1991) *349*:463; Harosh, et al., *Nucleic Acids Res.*

(1991) *19*:6331; Koonin , et al., *J. Gen. Virol.* (1992) *73*:989; http://www.expasy.ch/www/linder/HELICASES_TEXT.html).

Dual specificity phosphatase (DSPc). Dual specificity phosphatases (DSPs) are Ser/Thr and Tyr protein phosphatases that comprise a tertiary fold highly similar to that of tyrosine-specific phosphatases, except for a "recognition" region connecting helix alpha1 to strand beta1. This tertiary fold may determine differences in substrate specific between VH-1 related dual specificity phosphatase (VHR), the protein tyrosine phosphatases (PTPs), and other DSPs. Phosphatases are important in the control of cell growth, proliferation, differentiation and transformation.

EF Hand (EFhand). One SEQ ID NOcorresponds to a polynucleotide encoding a member of the EF-hand protein family, a calcium binding domain shared by many calcium-binding proteins belonging to the same evolutionary family (Kawasaki *et al.*, *Protein. Prof.* (1995) *2*:305-490). The domain is a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain, with a calcium ion coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand).

Eukaryotic Aspartyl Proteases (asp). Several of the polynucleotides of the invention correspond to a sequence encoding a novel eukaryotic aspartyl protease. Aspartyl proteases, known as acid proteases, (EC 3.4.23.-) are a widely distributed family of proteolytic enzymes (Foltmann., *Essays Biochem.* (1981) *17*:52; Davies, *Annu. Rev. Biophys. Chem.* (1990) *19*:189; Rao, *et al.*, *Biochemistry* (1991) *30*:4663) known to exist in vertebrates, fungi, plants, retroviruses and some plant viruses. Aspartate proteases of eukaryotes are monomeric enzymes which consist of two domains. Each domain contains an active site centered on a catalytic aspartyl residue.

Fibronectin Type II collagen-binding domain (FntypeII). One SEQ ID NOcorresponds to a polynucleotide encoding a polypeptide having a type II fibronectin collagen binding domain. Fibronectin is a plasma protein that binds cell surfaces and various compounds including collagen, fibrin, heparin, DNA, and actin. The major part of the sequence of fibronectin consists of the repetition of three types of domains, called type I, II, and III (Skorstengaardet al., *Eur. J. Biochem.* (1986) *161*:441). The type II domain, which is duplicated in fibronectin, is approximately forty residues long, contains four conserved cysteines involved in disulfide bonds and is part of the collagen-binding region

of fibronectin. .

G-Protein Alpha Subunit (G-alpha). One SEQ ID NOcorresponds to a gene encoding a member of the G-protein alpha subunit family. G-proteins are a family of membrane-associated proteins that couple extracellularly-activated integral-membrane receptors to intracellular effectors, such as ion channels and enzymes that vary the concentration of second messenger molecules. G-proteins are composed of 3 subunits (alpha, beta and gamma) which, in the resting state, associate as a trimer at the inner face of the plasma membrane. The alpha subunit, which binds GTP and exhibits GTPase activity, is about 350-400 amino acids in length with a molecular weight in the range of 40-45 kDa. Seventeen distinct types of alpha subunit have been identified in mammals, and fall into 4 main groups on the basis of both sequence similarity and function: alpha-s, alpha-q, alpha-i and alpha-12 (Simon *et al.*, *Science* (1993) *252*:802). They are often N-terminally acylated, usually with myristate and/or palmitoylate, and these fatty acid modifications can be important for membrane association and high- affinity interactions with other proteins.

Helicases conserved C-terminal domain (helicase_C). Some SEQ ID NOSrepresent polynucleotides encoding novel members of the DEAD/H helicase family. The DEAD and DEAH families are described above.

Helix-Loop-Helix (HLH) DNA Binding Domain (HLH). One SEQ ID NO corresponds to a sequence encoding an HLH domain. The HLH domain, which normally spans about 40 to 50 amino acids, is present in a number of eukaryotic transcription factors. The HLH domain is formed of two amphipathic helices joined by a variable length linker region that forms a loop that mediates protein dimerization (Murre et al. *Cell* (1989) *56*:777-783). Basic HLH proteins (bHLH), which have an extra basic region of about 15 amino acid residues adjacent the HLH domain and specifically bind to DNA, include two groups: class A (ubiquitous) and class B (tissue-specific). bHLH family members bind variations of the E-box motif (CANNTG). The homo- or heterodimerization mediated by the HLH domain is independent of, but necessary for DNA binding, as two basic regions are required for DNA binding activity. The HLH proteins lacking the basic domain function as negative regulators since they form heterodimers, but fail to bind DNA.

Kinase Domain of Tors. The TOR profile is directed towards a lipid kinase protein family. This family is composed of large proteins with a lipid and protein kinase domain and characterized through their sensitivity to rapamycin (an antifungal compound). TOR proteins are involved in signal transduction downstream of PI3 kinase and many other

signals. TOR (also called FRAP, RAFT) plays a role in regulating protein synthesis and cell growth., and in yeast controls translation initiation and early G1 progression. See, *e.g.*, Barbet *et al. Mol Biol Cell.* (1996) *7*(1):25-42; Helliwell *et al. Genetics* (1998) *148*:99-112.

MAP kinase kinase (mkk). Some SEQ ID NOS represent members of the MAP kinase kinase (mkk) family. MAP kinases (MAPK) are involved in signal transduction, and are important in cell cycle and cell growth controls. The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast, invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates. MAPKKs are essential transducers through which signals must pass before reaching the nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) *79*:193-207; Nishida *et al.*, *Trends Biochem Sci* (1993) *18*:128-31; Ruderman *Curr Opin Cell Biol* (1993) *5*:207-13; Dhanasekaran *et al.*, *Oncogene* (1998) *17*:1447-55; Kiefer *et al.*, *Biochem Soc Trans* (1997) *25*:491-8; and Hill, *Cell Signal* (1996) *8*:533-44.

Neurotransmitter-Gated Ion-Channel (neur_chan). Several of the sequences correspond to a sequence encoding a neurotransmitter-gated ion channel. Neurotransmitter-gated ion-channels, which provide the molecular basis for rapid signal transmission at chemical synapses, are post-synaptic oligomeric transmembrane complexes that transiently form a ionic channel upon the binding of a specific neurotransmitter. Five types of neurotransmitter-gated receptors are known: 1) nicotinic acetylcholine receptor (AchR); 2) glycine receptor; 3) gamma-aminobutyric-acid (GABA) receptor; 4) serotonin 5HT3 receptor; and 5) glutamate receptor. All known sequences of subunits from neurotransmitter-gated ion-channels are structurally related, and are composed of a large extracellular glycosylated N-terminal ligand-binding domain, followed by three hydrophobic transmembrane regions that form the ionic channel, followed by an intracellular region of variable length. A fourth hydrophobic region is found at the C-terminal of the sequence.

Protein Kinase (protkinase). Several sequences represent polynucleotides encoding protein kinases, which catalyze phosphorylation of proteins in a variety of pathways, and are implicated in cancer. Eukaryotic protein kinases (Hanks, *et al.*, *FASEB J.* (1995) *9*:576; Hunter, *Meth. Enzymol.* (1991) *200*:3; Hanks, *et al.*, *Meth. Enzymol.* (1991) *200*:38; Hanks, *Curr. Opin. Struct. Biol.* (1991) *1*:369; Hanks *et al.*, *Science* (1988) *241*:42) belong to a

very extensive family of proteins that share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. The first region, located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, located in the central part of the catalytic domain, contains a conserved an aspartic acid residue that is important for the catalytic activity of the enzyme (Knighton, *et al.*, *Science* (1991) *253*:407).

The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the alignment in (Hanks, *et al.*, *FASEB J.* (1995) *9*:576) and covers the entire catalytic domain.

<u>Protein Tyrosine Phosphatase (Y_phosphatase) (PTPase)</u>. Some SEQ ID NOS represent polynucleotides encoding a tyrosine-specific protein phosphatase, a kinase that catalyzes the removal of a phosphate groups attached to a tyrosine residue (EC 3.1.3.48) (PTPase) (Fischer *et al.*, *Science* (1991) *253*:401; Charbonneau *et al.*, *Annu. Rev. Cell Biol.* (1992) *8*:463; Trowbridge *Biol. Chem.* (1991) *266*:23517; Tonks *et al.*, *Trends Biochem. Sci.* (1989) *14*:497; and Hunter, *Cell* (1989) *58*:1013). PTPases are important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). Structurally, all known receptor PTPases are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. PTPase domains consist of about 300 amino acids. Two conserved cysteines are absolutely required for activity, with a number of other conserved residues in the immediate vicinity also important for activity.

<u>RNA Recognition Motif (rrm).</u> Some SEQ ID NOScorrespond to sequence encoding an RNA recognition motif, also known as an RRM, RBD, or RNP domain. This domain, which is about 90 amino acids long, is contained in eukaryotic proteins that bind single-stranded RNA (Bandziulis et al. *Genes Dev.* (1989) *3*:431-437; Dreyfuss et al. *Trends Biochem. Sci.* (1988) *13*:86-91). Two regions within the RNA-binding domain are highly conserved: the first is a hydrophobic segment of six residues (which is called the RNP-2 motif), the second is an octapeptide motif (which is called RNP-1 or RNP-CS).

SH2 Domain (SH2). One SEQ ID NOcorresponds to a sequence encoding an SH2 domain. The Src homology 2 (SH2) domain includes an approximately 100 amino acid residue domain, which is conserved in the oncoproteins Src and Fps, as well as in many other intracellular signal-transducing proteins (Sadowski et al. *Mol. Cell. Biol.* (1986) *6*:4396-4408; Russel et al. *FEBS Lett.* (1992) *304*:15-20). SH2 domains function as regulatory modules of intracellular signaling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and strictly phosphorylation-dependent manner. The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets (Kuriyan et al. *Curr. Opin. Struct. Biol.* (1993) *3*:828-837).

Thioredoxin family active site (Thioredox). One SEQ ID NO represents a polynucleotide encoding a protein of the thioredoxin family. Thioredoxins are small proteins of approximately one hundred amino acid residues that participate in various redox reactions via the reversible oxidation of an active center disulfide bond (Holmgren, *Annu. Rev. Biochem.* (1985) *54*:237; Gleason, et al., *FEMS Microbiol. Rev.* (1988) *54*:271; Holmgren A. *J. Biol. Chem.* (1989) *264*:13963; Eklund, et al. *Proteins* (1991) *11*:13). Thioredoxins exist in either reduced or oxidized forms where the two cysteine residues are linked in an intramolecular disulfide bond. The sequence around the redox-active disulfide bond is well conserved.

Trypsin (trypsin). Some SEQ ID NOS correspond to novel serine proteases of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved (Brenner *Nature* (1988) *334*:528). All sequences known to belong to this family are detected by the above consensus sequences, except for 18 different proteases which have lost the first conserved glycine. If a protein includes both the serine and the histidine active site signatures, the probability of it being a trypsin family serine protease is 100%.

WD Domain, G-Beta Repeats (WD_domain). One SEQ ID NOrepresents a member of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by

transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) *56*:615). The alpha subunit binds to and hydrolyzes GTP; the beta and gamma subunits are required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta has eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat).

wnt Family of Developmental Signaling Proteins (Wnt_dev_sign). Several of the sequences correspond to novel members of the wnt family of developmental signaling proteins. Wnt-1 (previously known as int-1), the seminal member of this family, (Nusse, *Trends Genet.* (1988) *4*:291) plays a role in intercellular communication and is important in central nervous system development. All wnt family proteins share the following features characteristic of secretory proteins: a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that may be involved in disulfide bonds. Wnt proteins generally adhere to the plasma membrane of secreting cells and are therefore likely to signal over only few cell diameters.

Zinc Finger, C2H2 Type (Zincfing_C2H2). Some SEQ ID NOS correspond to polynucleotides encoding members of the C2H2 type zinc finger protein family, which contain zinc finger domains that facilitate nucleic acid binding (Klug *et al.*, *Trends Biochem. Sci.* (1987) *12*:464; Evans *et al.*, *Cell* (1988) *52*:1; Payre *et al.*, *FEBS Lett.* (1988) *234*:245; Miller *et al.*, *EMBO J.* (1985) *4*:1609; and Berg, *Proc. Natl. Acad. Sci. USA* (1988) *85*:99). In addition to the conserved zinc ligand residues, a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld *et al.*, *J. Biomol. Struct. Dyn.* (1993) *11*:557) The best conserved position, which is generally an aromatic or aliphatic residue, is located four residues after the second cysteine.

Example 30: Differential Expression of Polynucleotides of the Invention: Description of Libraries and Detection of Differential Expression

The relative expression levels of the polynucleotides of the invention was assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 43 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepared the cDNA library, the

"nickname" of the library that is used in the tables below (in quotes), and the approximate number of clones in the library.

Table 43. Description of cDNA Libraries

Library (lib #)	Description	Number of Clones in Cluster
1	Km12 L4 Human Colon Cell Line, High Metastatic Potential (derived from Km12C); "High Met Colon"	307133
2	Km12C Human Colon Cell Line, Low Metastatic Potential; "Low Met Colon"	284755
3	MDA-MB-231 Human Breast Cancer Cell Line, High Metastatic Potential; micro-metastases in lung; "High Met Breast"	326937
4	MCF7 Human Breast Cancer Cell, Non Metastatic; "Low Met Breast"	318979
8	MV-522 Human Lung Cancer Cell Line, High Metastatic Potential; "High Met Lung"	223620
9	UCP-3 Human Lung Cancer Cell Line, Low Metastatic Potential; "Low Met Lung"	312503
12	Human microvascular endothelial cells (HMEC) – Untreated PCR (OligodT) cDNA library; "HMEC"	41938
13	Human microvascular endothelial cells (HMEC) – Basic fibroblast growth factor (bFGF) treated PCR (OligodT) cDNA library; "HMEC-bFGF"	42100
14	Human microvascular endothelial cells (HMEC) – Vascular endothelial growth factor (VEGF) treated PCR (OligodT) cDNA library; "HMEC-VEGF"	42825
15	Normal Colon – UC#2 Patient PCR (OligodT) cDNA library; "Normal Colon Tissue"	282722
16	Colon Tumor – UC#2 Patient PCR (OligodT) cDNA library; "Normal Colon Tumor Tissue"	298831
17	Liver Metastasis from Colon Tumor of UC#2 Patient PCR (OligodT) cDNA library; "High Met Colon Tissue"	303467
18	Normal Colon – UC#3 Patient PCR (OligodT) cDNA library; "Normal Colon Tissue"	36216
19	Colon Tumor – UC#3 Patient PCR (OligodT) cDNA library; "Colon Tumor Tissue"	41388
20	Liver Metastasis from Colon Tumor of UC#3 Patient PCR (OligodT) cDNA library; "High Met Colon Tissue"	30956
21	GRRpz Human Prostate Cell Line; "Normal Prostate"	164801
22	Woca Human Prostate Cancer Cell Line; "Prostate Cancer"	162088

The KM12L4, KM12C, and MDA-MB-231 cell lines are described above. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, *e.g.*, Chandrasekaran *et al.*, *Cancer Res.* (1979) *39*:870 (MDA-MB-231 and MCF-7); Gastpar *et al.*, *J Med Chem* (1998) *41*:4965 (MDA-MB-231 and MCF-7); Ranson *et al.*, *Br J Cancer* (1998) *77*:1586 (MDA-MB-231 and MCF-7); Kuang *et al.*, *Nucleic Acids Res* (1998) *26*:1116 (MDA-MB-231 and MCF-7); Varki *et al.*, *Int J Cancer* (1987) *40*:46 (UCP-3); Varki *et al.*, *Tumour Biol.* (1990) *11*:327; (MV-522 and UCP-3); Varki *et al.*, *Anticancer Res.* (1990) *10*:637; (MV-522); Kelner *et al.*, *Anticancer Res* (1995) *15*:867 (MV-522); and Zhang *et al.*, *Anticancer Drugs* (1997) *8*:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-treated HMEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMEC were prepared by incubation with 20ng/ml VEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation. The GRRpz and WOca cell lines were provided by Dr. Donna M. Peehl, Department of Medicine, Stanford University School of Medicine. GRRpz was derived from normal prostate epithelium. The WOca cell line is a Gleason Grade 4 cell line.

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters. The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac *et al.*, *Genomics* (1996) *37*(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought

together computationally. These groups of clones are termed “clusters”. Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the “purity” of each cluster can be controlled. For example, artifacts of clustering may occur in computational clustering just as artifacts can occur in “wet-lab” screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1^{st}), and the determining the number of cDNA clones corresponding to the selected cluster in the second library (Clones in 2^{nd}). Differential expression of the selected cluster in the first library relative to the second library is expressed as a "ratio" of percent expression between the two libraries. In general, the "ratio" is calculated by: 1) calculating the percent expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the "number of clones" corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the "depth" of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5 , where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar, Biostatistical Analysis, Prentice Hall, Inc., USA, “Differences between Proportions,” pp 296-298 (1974).

Examples 31-38: Differential Expression of Polynucleotides of the Invention

A number of polynucleotide sequences have been identified that are differentially expressed between, for example, cells derived from high metastatic potential cancer tissue and low metastatic cancer cells, and between cells derived from high metastatic potential cancer tissue and normal tissue. Evaluation of the levels of expression of the genes corresponding to these sequences can be valuable in diagnosis, prognosis, and/or treatment (*e.g.*, to facilitate rationale design of therapy, monitoring during and after therapy, *etc.*). Moreover, the genes corresponding to differentially expressed sequences described herein can be therapeutic targets due to their involvement in regulation (*e.g.*, inhibition or promotion) of development of, for example, the metastatic phenotype. For example, sequences that correspond to genes that are increased in expression in high metastatic potential cells relative to normal or non-metastatic tumor cells may encode genes or regulatory sequences involved in processes such as angiogenesis, differentiation, cell replication, and metastasis.

Detection of the relative expression levels of differentially expressed polynucleotides described herein can provide valuable information to guide the clinician in the choice of therapy. For example, a patient sample exhibiting an expression level of one or more of these polynucleotides that corresponds to a gene that is increased in expression in metastatic or high metastatic potential cells may warrant more aggressive treatment for the patient. In contrast, detection of expression levels of a polynucleotide sequence that corresponds to expression levels associated with that of low metastatic potential cells may warrant a more positive prognosis than the gross pathology would suggest.

A number of polynucleotide sequences of the present invention are differentially expressed between human microvascular endothelial cells (HMEC) that have been treated with growth factors relative to untreated HMEC. Sequences that are differentially expressed between growth factor-treated HMEC and untreated HMEC can represent sequences encoding gene products involved in angiogenesis, metastasis (cell migration), and other development and oncogenic processes. For example, sequences that are more highly expressed in HMEC treated with growth factors (such as bFGF or VEGF) relative to untreated HMEC can serve as markers of cancer cells of higher metastatic potential. Detection of expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk

assessment for a patient. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The differential expression of the polynucleotides described herein can thus be used as, for example, diagnostic markers, prognostic markers, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers. The following examples provide relative expression levels of polynucleotides from specified cell lines and patient tissue samples.

Example 31: High Metastatic Potential Breast Cancer Versus Low Metastatic Breast Cancer Cells

The following tables summarize polynucleotides that represent genes that are differentially expressed between high metastatic potential and low metastatic potential breast cancer cells.

Table 44. High metastatic potential breast (lib3) > low metastatic potential (lib4) breast cancer cells

SEQ ID NO:	Lib3 Clones	Lib4 Clones	Lib3/Lib4
7309	40	0	39
7634	60	3	20
7562	14	0	14
7452	10	0	10
7479	10	1	10
7254	10	1	10
6537	10	1	10
7434	10	0	10
7522	19	2	9
7643	9	1	9
7409	8	1	8
6937	8	1	8
7630	8	0	8
7599	8	0	8
6925	8	1	8
7504	8	0	8
7543	7	0	7
7485	7	0	7
6452	7	0	7
7588	7	0	7
7639	22	3	7

SEQ ID NO:	Lib3 Clones	Lib4 Clones	Lib3/Lib4
6895	7	0	7
7533	6	0	6
7347	6	0	6
7068	18	3	6
7578	6	0	6
7395	6	0	6
6205	24	4	6
7654	6	0	6
7451	6	0	6
7644	11	2	5
6346	10	2	5
7015	26	6	4
6454	36	12	3
7621	75	28	3
7253	49	17	3

Table 45. Low metastatic potential breast (lib4) > high metastatic potential breast cancer cells (lib3)

SEQ ID NO:	Lib3 Clones	Lib4 Clones	Lib4/Lib3
6344	0	58	59
6822	1	23	24
6110	1	19	19
6795	0	14	14
6859	1	14	14
6116	1	13	13
6175	1	13	13
6811	0	10	10
7087	0	8	8
7295	0	8	8
6803	0	7	7
7224	4	26	7
6987	0	6	6
7242	2	11	6
6827	7	44	6
7614	3	15	5
6436	3	13	4
7045	4	13	3
7343	7	18	3
7281	497	1216	3

Example 32: High Metastatic Potential Lung Cancer Versus Low Metastatic Lung Cancer Cells

The following summarizes polynucleotides that represent genes differentially expressed between high metastatic potential lung cancer cells and low metastatic potential lung cancer cells:

Table 46. High metastatic potential lung (lib8) > low metastatic potential lung (lib9) lung cancer cells

SEQ ID NO:	Lib8 Clones	Lib9 Clones	Lib8/Lib9
6246	31	0	43
6747	43	2	30
7394	14	1	20
6153	11	0	15
6721	7	0	10
7418	7	1	10
6132	7	0	10
6717	18	3	8
6311	6	1	8
6657	19	4	7
6343	5	0	7
6295	5	0	7
7094	5	0	7
6598	5	0	7
7478	8	2	6
7277	17	4	6
7405	8	2	6
7253	15	4	5
7356	14	5	4
7281	710	266	4
7621	21	10	3

Table 47. Low metastatic potential lung (lib9) > high metastatic potential lung (lib8) cancer cells

SEQ ID NO:	Lib8 Clones	Lib9 Clones	Lib9/Lib8
7020	1	13	9
6918	1	13	9
6824	1	12	9
6437	1	12	9
7623	3	31	7
6794	4	26	5
7045	2	15	5
6840	3	23	5

SEQ ID NO:	Lib8 Clones	Lib9 Clones	Lib9/Lib 8
7069	8	27	2

Example 33: High Metastatic Potential Colon Cancer Versus Low Metastatic Colon Cancer Cells

Tables 48 and 49 summarize polynucleotides that represent genes differentially expressed between high metastatic potential and low metastatic potential colon cancer cells:

Table 48. High metastatic potential (lib1) > low metastatic potential (lib2) colon cancer cells

SEQ ID NO:	Lib1 Clones	Lib2 Clones	Lib1/Lib 2
6344	67	2	31
6183	12	0	11
6794	11	0	10
6153	13	3	4
7020	24	10	2
7345	24	9	2

Table 49. Low metastatic potential (lib2) > high metastatic potential colon cancer (lib1) cells

SEQ ID NO:	Lib1 Clones	Lib2 Clones	Lib2/Lib 1
7364	1	17	18
7210	0	15	16
7128	1	14	15
6205	5	60	13
7069	1	11	12
6187	1	11	12
7078	0	9	10
7363	3	28	10
6189	1	8	9
7652	1	8	9
7347	0	8	9
7302	2	17	9
6908	0	8	9
7350	0	7	8
7316	0	7	8
6862	0	7	8
7252	0	7	8
7103	0	7	8

SEQ ID NO:	Lib1 Clones	Lib2 Clones	Lib2/Lib 1
7077	0	7	8
6858	0	7	8
6972	0	6	6
7330	2	11	6
7279	0	6	6
7140	2	12	6
6881	0	6	6
7165	3	17	6
6866	0	6	6
6874	0	6	6
6888	0	6	6
6918	2	10	5
7354	7	23	4
7320	7	17	3
7080	8	19	3
6937	10	28	3
6435	14	34	3
7309	11	29	3
7297	5	14	3
7288	22	48	2

Example 34: High Metastatic Potential Colon Cancer Patient Tissue Vs. Normal Patient Tissue

Table 50 summarizes polynucleotides that represent genes differentially expressed between high metastatic potential colon cancer cells and normal colon cells of patient tissue. :

Table 50. High metastatic potential colon tissue (lib17) vs. normal colon tissue (lib15)

SEQ ID NO:	Lib15 Clones	Lib17 Clones	Lib17/Lib1 5
7518	1	13	12
7228	1	10	9
6826	1	9	8
7407	0	7	7
6174	9	48	5
6918	5	20	4
SEQ ID NO:	**Lib15 Clones**	**Lib17 Clones**	**Lib15/Lib1 7**
6559	8	1	9

Example 35: High Tumor Potential Colon Tissue Vs. Metastasized Colon Cancer Tissue

The following table summarizes polynucleotides that represent genes differentially expressed between high tumor potential colon cancer cels and cells derived from high metastatic potential colon cancer cells of a patient.

Table 51. High tumor potential colon tissue (lib16) vs. high metastatic colon tissue (lib17)

SEQ ID NO:	Lib16 Clones	Lib17 Clones	Lib16/Lib17
7281	14	4	4
SEQ ID NO:	**Lib16 Clones**	**Lib17 Clones**	**Lib17/Lib16**
6918	2	20	10

Example 36: High Tumor Potential Colon Cancer Patient Tissue Versus Normal Patient Tissue

Tables 13 and 14 summarize polynucleotides that represent genes differentially expressed between high metastatic potential colon cancer cells and normal colon cells in patient tissue:

Table 52. Higher expression in tumor potential colon tissue (lib16) vs. normal colon tissue (lib15)

SEQ ID NO:	Lib15 Clones	Lib16 Clones	Lib16/Lib15
7407	0	8	8
6174	9	28	3

Table 53. Higher expression in normal colon tissue (lib15) vs. tumor potential colon tissue (lib16)

SEQ ID NO:	Lib15 Clones	Lib16 Clones	Lib15/Lib16
6559	8	0	8
7195	12	3	4

Example 37: Growth Factor-Stimulated Human Microvascular Endothelial Cells (HMEC) Relative to Untreated HMEC

The following tables summarize polynucleotides that represent genes differentially expressed between growth factor-treated and untreated HMEC.

Table 54. Higher expression in bFGF treated HMEC (lib13) vs. untreated HMEC (lib12)

SEQ ID NO:	Lib12 Clones	Lib13 Clones	Lib13/Lib12
7616	9	23	3
7634	17	35	2

Table 55. Higher expression in VEGF treated HMEC (lib14) vs. untreated HMEC (lib12)

SEQ ID NO:	Lib12 Clones	Lib14 Clones	Lib14/Lib12
7250	2	12	6
7322	2	10	5
7634	17	38	2

Example 38: Polynucleotides Differentially Expressed in Human Prostate Cancer Cells Relative to Normal Human Prostate Cells

The following tables summarize identified polynucleotides that represent genes differentially expressed between prostate cancer cells and normal prostate cells:

Table 56. Higher expression in normal prostate cells (lib21) relative to prostate cancer cells (lib22)

SEQ ID NO:	Lib21 Clones	Lib22 Clones	Lib21/Lib22
7621	6	0	6
6344	116	51	2
7299	22	9	2

Table 57 Higher expression in prostate cancer cells (lib22) relative to normal prostate cells (lib21)

SEQ ID NO:	Lib21 Clones	Lib22 Clones	Lib22/Lib21
7309	0	34	35
6436	1	12	12
6795	0	11	11

Example 39: Differential Expression Across Multiple Libraries

A number of polynucleotide sequences have been identified that represent genes that are differentially expressed across multiple libraries. Expression of these sequences in a tissue or any origin can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. These polynucleotides can also serve as non-tissue specific markers of, for example, risk of metastasis of a tumor. Table 58 summarizes this data.

Table 58. Genes Differentially Expressed Across Multiple Library Comparisons

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	Ratio
6153	High Met Lung (lib8) > Low Met Lung (lib9)	15
6153	High Met Colon (lib1) > Low Met Colon (lib2)	4
6174	High Met Colon Tissue (lib17) > Normal Colon Tissue (lib15)	5
6174	Normal Colon Tumor Tissue (lib16) > Normal Colon Tissue (lib15)	3
6205	High Met Breast (lib3) > Low Met Breast (lib4)	6
6205	Low Met Colon (lib2) > High Met Colon (lib1)	13
6344	High Met Colon (lib1) > Low Met Colon (lib2)	31
6344	Normal Prostate (lib21) > Prostate Cancer (lib22)	2
6344	Low Met Breast (lib4) > High Met Breast (lib3)	59
6436	Prostate Cancer (lib22) > Normal Prostate (lib21)	12
6436	Low Met Breast (lib4) > High Met Breast (lib3)	4
6559	Normal Colon Tissue (lib15) > High Met Colon Tissue (lib17)	9
6559	Normal Colon Tissue (lib15) > Normal Colon Tumor Tissue (lib16)	8
6794	High Met Colon (lib1) > Low Met Colon (lib2)	10
6794	Low Met Lung (lib9) > High Met Lung (lib8)	5
6795	Low Met Breast (lib4) > High Met Breast (lib3)	14
6795	Prostate Cancer (lib22) > Normal Prostate (lib21)	11
6918	High Met Colon Tissue (lib17) > Normal Colon Tumor Tissue (lib16)	10
6918	Low Met Lung (lib9) > High Met Lung (lib8)	9
6918	Low Met Colon (lib2) > High Met Colon (lib1)	5
6918	High Met Colon Tissue (lib17) > Normal Colon Tissue (lib15)	4
6937	High Met Breast (lib3) > Low Met Breast (lib4)	8
6937	Low Met Colon (lib2) > High Met Colon (lib1)	3
7020	High Met Colon (lib1) > Low Met Colon (lib2)	2
7020	Low Met Lung (lib9) > High Met Lung (lib8)	9
7045	Low Met Lung (lib9) > High Met Lung (lib8)	5
7045	Low Met Breast (lib4) > High Met Breast (lib3)	3
7069	Low Met Colon (lib2) > High Met Colon (lib1)	12

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	Ratio
7069	Low Met Lung (lib9) > High Met Lung (lib8)	2
7253	High Met Lung (lib8) > Low Met Lung (lib9)	5
7253	High Met Breast (lib3) > Low Met Breast (lib4)	3
7281	Normal Colon Tumor Tissue (lib16) > High Met Colon Tissue (lib17)	4
7281	High Met Lung (lib8) > Low Met Lung (lib9)	4
7281	Low Met Breast (lib4) > High Met Breast (lib3)	3
7309	High Met Breast (lib3) > Low Met Breast (lib4)	39
7309	Prostate Cancer (lib22) > Normal Prostate (lib21)	35
7309	Low Met Colon (lib2) > High Met Colon (lib1)	3
7347	High Met Breast (lib3) > Low Met Breast (lib4)	6
7347	Low Met Colon (lib2) > High Met Colon (lib1)	9
7407	Normal Colon Tumor Tissue (lib16) > Normal Colon Tissue (lib15)	8
7407	High Met Colon Tissue (lib17) > Normal Colon Tissue (lib15)	7
7621	Normal Prostate (lib21) > Prostate Cancer (lib22)	6
7621	High Met Lung (lib8) > Low Met Lung (lib9)	3
7621	High Met Breast (lib3) > Low Met Breast (lib4)	3
7634	High Met Breast (lib3) > Low Met Breast (lib4)	20
7634	HMEC-VEGF (lib14) > HMEC (lib12)	2
7634	HMEC-bFGF (lib13) > HMEC (lib12)	2

Key for Table 58: High Met = high metastatic potential; Low Met = low metastatic potential; met = metastasized; tumor = non-metastasized tumor; HMEC = human microvascular endothelial cell; bFGF = bFGF treated; VEGF = VEGF treated.

Example 40: Identification of Contiguous Sequences Having a Polynucleotide of the Invention

The novel polynucleotides were used to screen publicly available and proprietary databases to determine if any of the polynucleotides of SEQ ID NOS:8707-8803 would facilitate identification of a contiguous sequence, *e.g.*, the polynucleotides would provide sequence that would result in 5' extension of another DNA sequence, resulting in production of a longer contiguous sequence composed of the provided polynucleotide and the other DNA sequence(s). Contiging was performed using the Gelmerge application (default settings) of GCG from the Univ. of Wisconsin.

Using these parameters, 97 contiged sequences were generated. These contiged sequences are provided as SEQ ID NOS: 8707-8803 (see Table 41C). Table 41C provides the SEQ ID NO of the contig sequence, the name of the sequence used to create the contig, and the accession number of the publicly available tentative human consensus (THC)

sequence used with the sequence of the corresponding sequence name to provide the contig. The sequence name of Table 41C can be correlated with the SEQ ID NO: of the polynucleotide of the invention using Tables 41A and 41B.

Table 41C		
SEQ ID NO:	Sequence Name	THC Accession No.
8707	RTA00000587F.p.24.1.Seq	THC226834
8708	RTA00000629F.l.02.1.Seq	THC210324
8709	RTA00000623F.n.17.1.Seq	THC208388
8710	RTA00000593F.i.08.2.Seq	H91190
8711	RTA00000622F.b.03.1.Seq	AA554045
8712	RTA00000618F.e.06.1.Seq	THC226692
8713	RTA00000592F.o.02.1.Seq	AA099789
8714	RTA00000618F.c.04.1.Seq	THC222808
8715	RTA00000590F.i.01.1.Seq	THC173163
8716	RTA00000606F.o.14.1.Seq	THC223717
8717	RTA00000626F.d.07.1.Seq	THC234888
8718	RTA00000587F.l.08.1.Seq	THC104384
8719	RTA00000586F.a.13.1.Seq	THC140691
8720	RTA00000617F.a.17.1.Seq	THC221850
8721	RTA00000615F.b.23.1.Seq	THC205191
8722	RTA00000632F.f.10.1.Seq	N39216
8723	RTA00000607F.o.13.2.Seq	THC233619
8724	RTA00000622F.c.12.1.Seq	THC118482
8725	RTA00000625F.b.07.1.Seq	THC223154
8726	RTA00000587F.j.01.1.Seq	H63018
8727	RTA00000608F.i.15.1.Seq	THC216448
8728	RTA00000592F.j.06.1.Seq	THC148215
8729	RTA00000589F.b.14.1.Seq	THC158020
8730	RTA00000633F.g.19.1.Seq	THC202541
8731	RTA00000620F.o.07.1.Seq	THC155200
8732	RTA00000586F.p.01.1.Seq	AA558590
8733	RTA00000630F.l.10.1.Seq	THC204748
8734	RTA00000626F.c.13.1.Seq	AA159259
8735	RTA00000591F.m.06.1.Seq	THC227858
8736	RTA00000630F.i.11.1.Seq	THC228806
8737	RTA00000621F.h.08.1.Seq	THC163604
8738	RTA00000589F.d.10.1.Seq	THC177076
8739	RTA00000597F.p.01.1.Seq	THC210746
8740	RTA00000619F.c.13.1.Seq	R57955
8741	RTA00000607F.c.07.2.Seq	THC208762

8742	RTA00000595F.b.02.1.Seq	THC233682
8743	RTA00000631F.h.04.1.Seq	THC223281
8744	RTA00000596F.p.18.1.Seq	THC197103
8745	RTA00000586F.o.13.1.Seq	THC222729
8746	RTA00000610F.p.17.1.Seq	EST19015
8747	RTA00000596F.c.05.1.Seq	EST72617
8748	RTA00000632F.j.19.1.Seq	THC90741
8749	RTA00000607F.e.23.2.Seq	AA639216
8750	RTA00000628F.b.19.1.Seq	THC118075
8751	RTA00000609F.d.13.1.Seq	THC195579
8752	RTA00000621F.k.03.1.Seq	EST70278
8753	RTA00000592F.l.04.1.Seq	THC91941
8754	RTA00000592F.k.09.1.Seq	THC229803
8755	RTA00000622F.e.17.1.Seq	R57425
8756	RTA00000628F.g.13.1.Seq	THC176706
8757	RTA00000592F.k.23.1.Seq	THC232202
8758	RTA00000609F.m.04.2.Seq	AA507611
8759	RTA00000626F.b.04.1.Seq	EST69420
8760	RTA00000591F.m.01.1.Seq	H41850
8761	RTA00000608F.n.23.1.Seq	THC214886
8762	RTA00000583F.d.19.1.Seq	THC229251
8763	RTA00000621F.p.15.1.Seq	THC212450
8764	RTA00000583F.n.05.1.Seq	AA252468
8765	RTA00000597F.f.17.1.Seq	THC219322
8766	RTA00000606F.l.10.1.Seq	THC225232
8767	RTA00000618F.n.14.1.Seq	THC216591
8768	RTA00000612F.h.05.3.Seq	THC158250
8769	RTA00000619F.a.24.1.Seq	AA437370
8770	RTA00000617F.k.13.1.Seq	AA244445
8771	RTA00000623F.h.07.1.Seq	THC212330
8772	RTA00000620F.e.01.1.Seq	THC167493
8773	RTA00000620F.h.10.1.Seq	THC232456
8774	RTA00000589F.e.21.2.Seq	THC208239
8775	RTA00000626F.b.22.1.Seq	THC225644
8776	RTA00000620F.i.16.1.Seq	AA536090
8777	RTA00000613F.c.17.1.Seq	THC92470
8778	RTA00000621F.c.12.1.Seq	THC156244
8779	RTA00000618F.b.17.1.Seq	THC209838
8780	RTA00000585F.d.16.1.Seq	THC211870
8781	RTA00000592F.a.06.1.Seq	THC233200
8782	RTA00000583F.p.08.1.Seq	THC196844
8783	RTA00000622F.h.21.1.Seq	EST12698
8784	RTA00000591F.h.03.1.Seq	THC213771

8785	RTA00000620F.g.22.1.Seq	THC224063
8786	RTA00000588F.l.20.2.Seq	R84876
8787	RTA00000614F.a.20.1.Seq	R84876
8788	RTA00000611F.n.14.3.Seq	THC200742
8789	RTA00000619F.f.23.1.Seq	THC227573
8790	RTA00000608F.g.24.1.Seq	T93977
8791	RTA00000595F.o.01.2.Seq	EST61392
8792	RTA00000608F.b.23.1.Seq	THC161665
8793	RTA00000606F.o.23.1.Seq	AA464645
8794	RTA00000588F.i.22.3.Seq	THC162216
8795	RTA00000610F.i.13.1.Seq	AA595068
8796	RTA00000608F.b.15.1.Seq	EST11866
8797	RTA00000597F.e.16.1.Seq	N88730
8798	RTA00000610F.h.13.1.Seq	THC195895
8799	RTA00000611F.h.21.2.Seq	EST46722
8800	RTA00000584F.b.06.1.Seq	EST02998
8801	RTA00000584F.b.06.2.Seq	EST02998
8802	RTA00000608F.j.05.1.Seq	EST60433
8803	RTA00000588F.b.03.1.Seq	THC164651

The contiged sequences (SEQ ID NOS: 8707-8803) thus represent longer sequences that encompass a polynucleotide sequence of the invention. The contiged sequences were then translated in all three reading frames to determine the best alignment with individual sequences using the BLAST programs as described above. The sequences were masked using the XBLAST program for masking low complexity as described above in Example 27. Several of the contiged sequences were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 42B, inserted prior to claims). Thus the invention encompasses fragments, fusions, and variants of such polynucleotides that retain biological activity associated with the protein family and/or functional domain identified herein.

Descriptions of the profiles for the indicated protein families and functional domains are provided 3 above. A description of the profile for PR55 is provided below.

Protein Phosphatase 2A Regulatory Subunit PR55 (PR55). Several of the contigs correspond to a sequence encoding a protein comprising a protein phosphatase 2A (PP2A) regulatory subunit PR55. PP2A is a serine/threonine phosphatase involved in many aspects of cellular function including the regulation of metabolic enzymes and proteins involved in signal transduction. PP2A is a trimeric enzyme comprising a core composed of a catalytic

subunit associated with a 65 Kd regulatory subunit (PR65, also called subunit A). This complex associates with a third variable subunit (subunit B), which confers distinct properties to the holoenzyme (Mayer-Jaekel et al. *Trends Cell Biol.* (1994) *4*:287-291). One of the forms of the variable subunit is a 55 Kd protein (PR55) which is highly conserved in mammals and may facilitate substrate recognition or targeting the enzyme complex to the appropriate subcellular compartment. The PR55 subunit comprises two conserved sequences of 15 residues; one located in the N-terminal region, the other in the center of the protein.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Deposit Information. The following materials were deposited with the American Type Culture Collection (CMCC = Chiron Master Culture Collection).

Table 59. Cell Lines Deposited with ATCC

Cell Line	**Deposit Date**	**ATCC Accession No.**	**CMCC Accession No.**
KM12L4-A	March 19, 1998	CRL-12496	11606
Km12C	May 15, 1998	CRL-12533	11611
MDA-MB-231	May 15, 1998	CRL-12532	10583
MCF-7	October 9, 1998	CRL-12584	10377

In addition, pools of selected clones, as well as libraries containing specific clones, were assigned an "ES" number (internal reference) and deposited with the ATCC. Table 60 below provides the ATCC Accession Nos. of the ES deposits, all of which were deposited on or before May 13, 1999. The names of the clones contained within each of these deposits are provided in the tables numbered 61 -63 (inserted before the claims).

Table 60: Pools of Clones and Libraries Deposited with ATCC on or before May 14, 1999

ES #	ATCC Accession #	ES #	ATCC Accession #	ES #	ATCC Accession #
34		41		48	
35		42		49	
36		43		50	
37		44		51	
38		45		52	
39		46		53	
40		47		54	

The deposits described herein are provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby.

Retrieval of Individual Clones from Deposit of Pooled Clones. Where the ATCC deposit is composed of a pool of cDNA clones or a library of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones in the pool or library were then deposited as a pool of equal mixtures in the composite deposit. Particular clones can be obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a T_m of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C).

Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

Table 61 Deposits of Pooled Clones

ES34	ES35	ES36	ES37
M00006992C:G02	M00005468A:D08	M00005452C:A02	M00022171D:B08
M00006756D:E10	M00021892B:H03	M00001382C:C09	M00008061A:F02
M00003984C:F04	M00001390A:C06	M00004841C:B09	M00003820C:A09
M00007125D:E03	M00022074D:F11	M00001441D:H05	M00022109B:A11
M00006650A:A10	M00005460B:D02	M00022716D:D08	M00005342D:F03
M00001452B:H06	M00022423B:D03	M00022828C:E04	M00022070B:C10
M00022972D:C10	M00007140A:F11	M00004350B:F06	M00006966B:B09
M00022305C:A01	M00004081B:C11	M00005685B:D08	M00022381C:C12
M00007010B:H01	M00005480A:H12	M00004190A:A09	M00003991B:B05
M00021946D:C11	M00008015D:E09	M00004054D:D02	M00022404D:G05

ES38	ES39	ES40	ES41
M00021912B:H11	M00007118B:B04	M00006993B:B09	M00007974B:C11
M00005378C:A10	M00007019A:B01	M00004242C:C01	M00021860B:G06
M00022578C:B07	M00021682B:D12	M00007986C:C05	M00006927C:F12
M00005513A:D08	M00005411D:A03	M00004115A:G09	M00022582C:E12
M00022176C:A08	M00006641C:H02	M00022600C:A06	M00006618C:G08
M00006822D:F07	M00007041B:C05	M00005384A:A01	M00005450B:B01
M00004031A:B04	M00005444B:E11	M00021667D:E03	M00001417B:E01
M00021927D:D12	M00022745B:G02	M00008078C:C06	M00003825B:A05
M00001553D:B06	M00022685A:F11	M00007985A:B09	M00001370B:B04
M00022404B:H05	M00004446A:G01	M00007953B:B03	M00006727B:E09

ES42	ES43	ES44	ES45
M00001478A:B06	M00006923B:H08	M00006615B:F05	M00005468D:F04

M00003972B:A11	M00005377D:F11	M00005486C:B03	M00006720C:C11
M00005477C:D08	M00006640B:H09	M00007124C:A11	M00005817D:E12
M00006745A:A01	M00005404C:F02	M00006995D:A03	M00001669B:A03
M00007090B:A02	M00004030A:G12	M00007149D:G06	M00003998A:G12
M00007152A:B04	M00006704D:D03	M00006990D:D06	M00004045A:B12
M00006953B:H10	M00006810D:A05	M00005530B:E04	M00004130D:E04
M00005399D:B02	M00005481C:A05	M00003918C:E07	M00004160A:D07
M00006987B:F04	M00005411A:C07	M00007163A:B10	M00001655A:F07
M00005772A:F03	M00003970A:G10	M00005485C:A03	M00001468D:D11

ES46
M00004217A:A05
M00004183D:B07
M00001415D:A05
M00004158C:F03
M00004031D:G02

Table 62. Library Deposits

ES47	**ES48**	**ES49**	**ES50**
M00001399D:F09	M00004217D:G10	M00004508A:G12	M00021653A:G07
M00001455A:C03	M00004218C:G10	M00004508B:G02	M00021654C:A02
M00001456C:F02	M00004252D:H08	M00001432B:H08	M00021660C:G04
M00001487D:G03	M00004253B:A10	M00001432C:G01	M00021665A:D04
M00001539B:B01	M00004253B:F06	M00003992D:G01	M00021670B:G11
M00001565A:A02	M00004253C:E10	M00005326B:F03	M00021678A:B08
M00001572C:E07	M00004260A:B07	M00005332A:H10	M00021680B:C01
M00001582D:B10	M00004260C:A12	M00005342A:C04	M00021681C:B10
M00001584C:A03	M00004260C:E10	M00005342A:D04	M00021690D:E05
M00001586A:F09	M00001339B:A03	M00005349B:G01	M00021692A:E03
M00001588D:H08	M00001342C:A04	M00005352B:D02	M00021692C:E06

M00001610B:A01	M00001344D:G11	M00005354C:E02	M00021694B:A07
M00001618B:F02	M00001345A:A12	M00005356A:D09	M00021698B:B12
M00001618C:E06	M00001347A:G06	M00005359D:G07	M00021828A:C08
M00001621C:A04	M00001347B:H01	M00005378A:A08	M00021841C:D07
M00001626B:H05	M00001353B:D11	M00005383D:D06	M00021859A:D04
M00001641B:G05	M00001355B:A01	M00005383D:E07	M00021861C:A02
M00001648C:F06	M00001358D:D09	M00005385C:G05	M00021862A:A04
M00001649D:H05	M00001359A:B07	M00005388D:F09	M00021862D:F01
M00001656D:F11	M00001362A:C10	M00005390B:G10	M00021886D:E04
M00001660A:F10	M00001362B:A09	M00005397C:B03	M00021897B:A06
M00001669A:H11	M00001365D:D12	M00005399A:D01	M00021905A:G05
M00003741A:E01	M00001365D:H09	M00005409D:C02	M00021905B:A01
M00003745C:E03	M00001370A:G09	M00005415C:G08	M00021906C:G11
M00003746A:E01	M00001370B:B12	M00005417A:E10	M00021910A:C10
M00003748B:B06	M00001374D:D09	M00005442D:C05	M00021927A:C11
M00003749B:C08	M00001376B:C11	M00005446A:G01	M00021927B:F01
M00003749D:G07	M00001377A:D03	M00005446C:D12	M00021932C:C05
M00003752A:B06	M00001377A:E01	M00005454C:H12	M00021932C:G10
M00003752D:D09	M00001377C:B08	M00005455A:D01	M00021947A:C01
M00003753C:B01	M00001387A:A04	M00005455A:G03	M00021952B:F11
M00003754C:F01	M00001387D:C07	M00005462C:B02	M00021954A:A03
M00003756C:C08	M00001389B:B06	M00005469D:C11	M00021964A:C04
M00003759A:E10	M00001390A:H01	M00005480C:B12	M00021967D:E08
M00003762A:D11	M00001399C:E10	M00005483D:A12	M00021977D:E02
M00003763B:D03	M00001401D:D04	M00005484A:D09	M00021978A:F08
M00003763D:F06	M00001402D:C07	M00005491B:C03	M00021982C:F08
M00003765D:E02	M00001402D:H03	M00005493B:C08	M00021983B:B03
M00003766B:G04	M00001403B:A01	M00005494D:F11	M00021983D:B10
M00003767C:F04	M00001405D:F05	M00005496C:A01	M00022005C:G03
M00003769B:A04	M00001406C:A11	M00005496D:A10	M00022032A:E07
M00003769D:G12	M00001406D:H01	M00005497B:H07	M00022049A:A02
M00003770D:C07	M00001407B:A08	M00005497C:C07	M00022049A:D06

M00003771A:G09	M00001407D:H11	M00005497C:C12	M00022054D:C05
M00003771D:A10	M00001411A:D01	M00005497C:E03	M00022064C:H07
M00003773A:C09	M00001411C:G02	M00005498B:F08	M00022067D:C05
M00003773B:E09	M00001412A:A11	M00005498C:G05	M00022068B:H11
M00003773B:G08	M00001415D:E12	M00005508B:B04	M00022068D:D12
M00003773C:G06	M00001417C:E02	M00005524C:B01	M00022069D:G02
M00003773D:C02	M00001421A:H07	M00005528D:A10	M00022071B:D05
M00003789C:E03	M00001422D:D02	M00005530B:D03	M00022071C:D09
M00003790B:F12	M00001423C:D06	M00005534B:H10	M00022075D:F05
M00003793C:D11	M00001424A:H09	M00005548B:E03	M00022081C:G11
M00003796B:C07	M00001425C:E10	M00005550B:D09	M00022084B:F04
M00003797D:H06	M00001426A:F09	M00005565C:A08	M00022085C:C04
M00003801D:F05	M00001426D:D09	M00005589C:B03	M00022090A:G08
M00003805A:G05	M00001431A:C10	M00005616B:D05	M00022093A:A05
M00003808C:D09	M00001431A:E05	M00005620C:C05	M00022093D:B10
M00003809A:A12	M00001432A:F12	M00005621A:G10	M00022094B:G10
M00003809A:H12	M00001432B:H08	M00005621D:F01	M00022106C:F04
M00003813D:A06	M00001432C:G01	M00005631A:A11	M00022110A:E04
M00003818A:F09	M00001433A:C07	M00005632C:D06	M00022114C:B02
M00003818B:A01	M00001434A:A01	M00005637B:D12	M00022117C:G07
M00003819D:G09	M00001435A:F03	M00005642B:C03	M00022128A:D04
M00003821C:E04	M00001435A:G01	M00005647D:D09	M00022139A:C01
M00003822A:G05	M00001435B:G10	M00005655B:C02	M00022149B:D05
M00003825C:B02	M00001435C:G08	M00005703A:C08	M00022150A:H06
M00003825C:B12	M00001435D:A06	M00005704A:B11	M00022153D:D11
M00003833B:A11	M00001436D:C10	M00005708D:B03	M00022157A:F12
M00003834A:A03	M00001437B:B05	M00005710A:C08	M00022157B:A10
M00003835D:H05	M00001438C:H05	M00005720A:D03	M00022169D:C02
M00003839D:G06	M00001439B:F10	M00005722D:G03	M00022170D:H09
M00003841A:E09	M00001439C:A01	M00005743B:F02	M00022175A:A11
M00003841B:D05	M00001439C:G06	M00005763B:H09	M00022176A:E08
M00003843A:B01	M00001442A:D08	M00005765C:C04	M00022178D:H01

M00003844C:D04	M00001443D:A01	M00005810C:D04	M00022183A:G03
M00003844C:H05	M00001444A:A09	M00005813D:F06	M00022189A:A01
M00003846B:H02	M00001446D:B10	M00005818C:E08	M00022198A:C12
M00003850B:D11	M00001452D:E05	M00005818C:G01	M00022199C:F03
M00003852D:D03	M00001453D:F09	M00006576D:F11	M00022202C:F11
M00003859C:B09	M00001463C:A01	M00006577B:H12	M00022206B:G06
M00003868D:F02	M00001466C:F02	M00006587A:H08	M00022212C:C02
M00003868D:F07	M00001471C:G03	M00006594A:E08	M00022216D:C01
M00003871A:E09	M00001488B:G12	M00006596D:H04	M00022218C:B06
M00003884D:A12	M00001489B:F08	M00006601C:A07	M00022218D:B12
M00003887B:C03	M00001489D:C08	M00006601C:E06	M00022220C:F08
M00003888B:A10	M00001490B:G04	M00006609A:G10	M00022221D:E08
M00003888C:E01	M00001491C:C01	M00006633C:E11	M00022226C:B06
M00003890B:H07	M00001496A:B03	M00006633D:A06	M00022226D:A07
M00003890D:C03	M00001496D:D02	M00006634B:C02	M00022231A:F12
M00003892D:D04	M00001500A:D09	M00006636A:B08	M00022231C:A04
M00003893C:D12	M00001504D:D09	M00006644A:B11	M00022236D:A03
M00003895D:A03	M00001505A:E09	M00006644D:C02	M00022239A:A10
M00003896B:F08	M00001506A:F01	M00006686A:G12	M00022239B:B07
M00003896D:B01	M00001517D:C03	M00006692B:E04	M00022239D:A07
M00003903C:H03	M00001518D:A10	M00006728D:G10	M00022252C:E06
M00003905C:B01	M00001536B:B11	M00006733D:G12	M00022253B:E06
M00003905C:E10	M00001537B:C12	M00006734A:H12	M00022254C:D08
M00003906C:H12	M00001542C:D10	M00006735A:H02	M00022255A:C08
M00003909D:G01	M00001542C:F06	M00006764B:D05	M00022255D:E03
M00003911C:G05	M00001543A:E04	M00006765B:H06	M00022258C:F06
M00003912B:G11	M00001546B:H01	M00006785B:F09	M00022259B:G02
M00003912C:C11	M00001551D:C12	M00006791B:B08	M00022278C:E03
M00003914C:E03	M00001552B:D01	M00006796A:C03	M00022278D:F10
M00003915A:D09	M00001556D:A11	M00006800C:G08	M00022288C:D04
M00003915C:G01	M00001557C:B08	M00006814A:F07	M00022289A:D05
M00003920B:A10	M00001558B:A12	M00006819A:D10	M00022289D:B06

M00003921D:C06	M00001560C:C01	M00006820A:G05	M00022294A:D11
M00003923A:H07	M00001561B:C10	M00006821C:C10	M00022296B:C11
M00003936C:F10	M00001597C:B03	M00006822A:D07	M00022305A:H11
M00003948B:B03	M00001623B:B01	M00006823D:D12	M00022364C:G12
M00003949B:A08	M00001623D:A09	M00006826B:H03	M00022366B:E09
M00003949B:D05	M00001644D:F09	M00006828D:C12	M00022372B:D03
M00003961B:A12	M00003784C:B09	M00006832D:F11	M00022381A:F05
M00003961C:G02	M00003785D:E01	M00006846A:B01	M00022382D:H11
M00003962B:B09	M00003862C:H10	M00006850C:D09	M00022386A:A07
M00003963B:D12	M00003864B:A04	M00006850C:G07	M00022386B:D11
M00003973A:C05	M00003864D:G05	M00006851C:H09	M00022386C:A04
M00003973B:H06	M00003992C:G01	M00006863B:E06	M00022386C:D07
M00003976D:D12	M00003992D:G01	M00006866C:F03	M00022399C:A10
M00003977C:A08	M00003994C:C11	M00006867C:E07	M00022407C:H11
M00003980B:F12	M00003996D:C04	M00006868D:E02	M00022411D:G09
M00003980C:G10	M00003997D:D07	M00006870C:H06	M00022412A:C08
M00003981C:E04	M00003998A:D03	M00006873B:G11	M00022444A:A11
M00003983C:E07	M00003998C:H10	M00006875A:A02	M00022449C:B01
M00003987D:F06	M00003999C:C12	M00006877B:E05	M00022452C:B03
M00004027A:B10	M00004046A:F04	M00006879A:H11	M00022457C:B01
M00004027C:H01	M00004051C:D02	M00006882A:D01	M00022495C:G05
M00004028C:B04	M00004052C:A08	M00006901D:A11	M00022504B:E03
M00004030B:B02	M00004052C:B05	M00006907C:D03	M00022505D:A12
M00004030B:C05	M00004054B:G02	M00006907D:C07	M00022509D:F06
M00004035D:E04	M00004054D:A03	M00006912B:E01	M00022527A:E05
M00004036B:F09	M00004055B:F06	M00006921B:E01	M00022527D:B03
M00004036C:D01	M00004058B:C11	M00006960D:E06	M00022531B:D07
M00004037A:A07	M00004058C:E08	M00006963A:H11	M00022535D:B11
M00004037B:B05	M00004059A:G09	M00006966C:B07	M00022535D:C04
M00004038C:C05	M00004060C:A02	M00006972A:F10	M00022536B:B04
M00004038C:D12	M00004060D:A07	M00006973C:E11	M00022551A:G03
M00004039D:D03	M00004063C:B11	M00006973D:E11	M00022556B:C04

M00004040B:B09	M00004143A:G12	M00006974B:F06	M00022556B:G02
M00004040C:G12	M00004143A:H07	M00006976C:E09	M00022562C:H10
M00004040D:B05	M00004145C:A03	M00007014C:B07	M00022578B:G05
M00004041B:F01	M00004146D:A07	M00007015C:G05	M00022578D:F03
M00004041D:E06	M00004147A:G03	M00007016C:E06	M00022583B:E05
M00004043D:C10	M00004149B:H12	M00007041B:G01	M00022587C:G04
M00004069D:G02	M00004153D:E06	M00007042A:E07	M00022594B:H12
M00004071A:H03	M00004154D:F11	M00007043A:B05	M00022598A:F11
M00004073D:B11	M00004159D:C04	M00007046A:D02	M00022599D:E07
M00004076D:B03	M00004166B:E10	M00007047B:D01	M00022604B:C11
M00004081C:A01	M00004166C:A03	M00007051D:D09	M00022607B:A04
M00004084C:G04	M00004166D:G07	M00007053B:H03	M00022613D:C04
M00004085B:G06	M00004196C:G05	M00007058A:C02	M00022651D:C06
M00004087C:F05	M00004234B:E03	M00007062A:D03	M00022666C:H11
M00004091A:E01	M00004234B:G06	M00007099A:F09	M00022681C:H02
M00004091B:C12	M00004236D:E07	M00007100C:D01	M00022682A:F12
M00004091B:G04	M00004236D:F04	M00007112B:C06	M00022698C:E06
M00004091C:F04	M00004240D:A07	M00007105D:C07	M00022701B:B12
M00004091D:D09	M00004242C:C02	M00007121A:A05	M00022708A:C08
M00004092A:C03	M00004244B:A02	M00007122A:G11	M00022708D:G10
M00004092A:D04	M00004245A:G09	M00007122B:A11	M00022725C:E09
M00004093D:D09	M00004245C:A03	M00007127B:A04	M00022726A:A06
M00004101D:A03	M00004247A:E01	M00007129A:G10	M00022730A:E04
M00004103B:C07	M00004247B:C11	M00007130B:B03	M00022737A:C08
M00004107C:A01	M00004248A:G08	M00007132D:G08	M00022763A:E10
M00004114C:F02	M00004263D:F06	M00007134C:F07	M00022824C:H11
M00004115A:F01	M00004272D:D02	M00007137D:C10	M00022835C:E06
M00004117B:F01	M00004273D:E11	M00007140D:C12	M00022854D:H07
M00004120A:C02	M00004277D:C08	M00007150A:C09	M00022856A:D02
M00004126B:G02	M00004281B:B05	M00007150A:H06	M00022856B:F04
M00004129A:H08	M00004283C:D03	M00007154A:E04	M00022856C:B11
M00004130C:A09	M00004285B:E01	M00007163A:F11	M00022893C:H11

M00004133D:A01	M00004297D:E08	M00007163B:A12	M00022897A:F04
M00004178B:F06	M00004298B:D04	M00007166B:E06	M00022900D:E08
M00004180B:F04	M00004308A:E06	M00007170D:A10	M00022900D:G03
M00004184B:F11	M00004324B:D09	M00007172A:A05	
M00004191B:G01	M00004328A:H06	M00007172D:C08	
M00004193A:C07	M00004329C:F11	M00007188A:D03	
M00004193C:H01	M00004331D:H08	M00007189D:A09	
M00004199D:C02	M00004332C:E09	M00007193D:A04	
M00004200A:A09	M00004337D:G08	M00007195B:B02	
M00004200A:G06	M00004345A:H06	M00007198C:A10	
M00004200D:A07	M00004383A:F02	M00007199D:B07	
M00004201D:C11	M00004385C:B11	M00007204C:F09	
M00004201D:E12	M00004388C:D05	M00007929B:H10	
M00004202B:A02	M00004406A:H03	M00007961A:B01	
M00004204A:D04	M00004408D:A10	M00007964B:D10	
M00004204A:D10	M00004410A:E03	M00007971A:B04	
M00004204B:A04	M00004412B:E03	M00007977C:E08	
M00004210A:B09	M00004421A:G04	M00007995D:E06	
M00004216D:E10	M00004447D:D10	M00008074D:C01	
M00004217A:A11	M00004460B:H09	M00008094A:E10	
	M00004465C:B10	M00021611D:D05	
	M00004465C:B12	M00021611D:H03	
	M00004467A:F09	M00021614B:G12	
	M00004467D:F09	M00021618D:D07	
	M00004491D:D07	M00021624A:D07	
	M00004497C:E09	M00021624B:A03	
	M00004501A:G06	M00021625A:C07	
	M00004506C:H10	M00021629D:D05	

Table 63 Library Deposits			
ES51	**ES52**	**ES53**	**ES54**
M00001448A:D05	M00001439B:E02	M00006621A:G10	M00021640A:G03

M00001458B:F06	M00001443A:E02	M00006626A:G11	M00021657B:C08
M00001530A:D11	M00001443D:C03	M00006629D:D04	M00021690B:B06
M00001563C:D06	M00001444A:G12	M00006630B:H06	M00021690C:B07
M00001564C:D04	M00001445B:E03	M00006631D:B02	M00022071C:C09
M00001569B:F04	M00001451B:H11	M00006631D:C04	M00022081C:B11
M00001575A:H02	M00001452B:F09	M00006631D:E09	M00022085C:A07
M00001589C:D12	M00001488B:H02	M00006635C:B10	M00022091B:B07
M00001589D:G10	M00001491D:E07	M00006636A:E06	M00022122D:D06
M00001590D:A07	M00001496C:H10	M00006636D:A05	M00022150D:D11
M00001598C:D10	M00001499A:D01	M00006636D:F11	M00022154A:C01
M00001599A:H09	M00001499A:D05	M00006640A:B01	M00022170D:H07
M00001609A:B12	M00001499B:H05	M00006640B:F05	M00022365A:A01
M00001614C:G04	M00001500B:H07	M00006640D:H08	M00022389B:H04
M00001626C:C10	M00001504C:H11	M00006641A:B03	M00022439A:E07
M00001634C:E12	M00001506D:A11	M00006643A:E10	M00022449D:F06
M00001639A:A04	M00001543A:D03	M00006644C:E09	M00022458B:E06
M00001640A:F02	M00001543A:F01	M00006648C:E04	M00022474A:H09
M00001640A:F04	M00001548C:A09	M00006650A:B11	M00022480B:E07
M00001647C:C07	M00001555D:F11	M00006656C:C10	M00022489C:A08
M00001649B:E08	M00001557B:D10	M00006664B:B04	M00022490C:A08
M00001654D:F06	M00001597A:C07	M00006664D:H09	M00022490C:C01
M00001658B:C07	M00001604B:D09	M00006665A:F07	M00022493C:B07
M00001659D:G08	M00001605D:G01	M00006665B:D10	M00022493C:C06
M00001663C:C03	M00001621D:B09	M00006674B:F04	M00022498C:C08
M00001675C:B03	M00001622C:F06	M00006676B:F11	M00022514A:D04
M00001677A:A06	M00001624A:A09	M00006676D:D11	M00022515D:C04
M00001677A:A12	M00001640D:C10	M00006679C:D07	M00022549B:G07
M00001678D:A12	M00001645B:C09	M00006681C:G04	M00022557B:A08
M00001679C:F03	M00003782D:F04	M00006695B:F08	M00022565C:H02
M00001681A:H09	M00003783C:A06	M00006698B:E06	M00022578D:A08
M00001687C:A06	M00003786D:C06	M00006699B:C07	M00022597B:F11
M00001693D:F07	M00003787B:D07	M00006705B:D02	M00022599A:C03

M00003746B:E12	M00003787D:A06	M00006712B:H10	M00022661B:E11
M00003766A:G09	M00003864C:D09	M00006717A:D04	M00022661D:H01
M00003795A:B01	M00003993A:E12	M00006721C:G07	M00022666B:E12
M00003796C:H03	M00003997B:H04	M00006725A:A03	M00022674D:G04
M00003797D:E10	M00003997D:G11	M00006725A:B03	M00022718D:G05
M00003799B:D02	M00004047B:G09	M00006727B:G08	M00022725C:B03
M00003809B:D08	M00004048D:A07	M00006728C:B06	M00022727B:C05
M00003811B:E07	M00004049D:G04	M00006737C:A08	M00022728A:A09
M00003812B:F08	M00004050A:F02	M00006738A:E05	M00022730D:E10
M00003812D:E08	M00004051C:D10	M00006739B:B10	M00022735B:B01
M00003815C:A06	M00004058B:F12	M00006739B:B12	M00022745A:B04
M00003815D:D01	M00004060C:A11	M00006739C:H07	M00022856B:D07
M00003816C:F10	M00004064A:B12	M00006743B:G12	M00022901D:C09
M00003818C:E09	M00004066A:E12	M00006744C:C06	M00022902D:D03
M00003819A:B09	M00004067C:D08	M00006745D:E08	M00022953B:C07
M00003819C:E04	M00004134A:F08	M00006751A:F03	M00022960D:E08
M00003820A:H04	M00004134A:H04	M00006758D:C01	M00022963A:D11
M00003820D:E02	M00004134C:B11	M00006760D:G12	M00022968A:F02
M00003824B:D06	M00004140B:B01	M00006763B:B11	M00022980B:E11
M00003825B:D12	M00004143C:F08	M00006769D:A04	M00022980C:A09
M00003826B:D01	M00004144D:B06	M00006770B:C05	M00022993A:F02
M00003829A:E02	M00004152C:E01	M00006771A:E06	M00023003C:A03
M00003832B:G03	M00004159D:H07	M00006771A:H07	M00023011A:A06
M00003833D:D06	M00004160A:A01	M00006771B:A09	M00023021A:H08
M00003835A:E03	M00004161B:A12	M00006771B:F03	M00023023A:B12
M00003837C:F05	M00004163A:D11	M00006774D:C01	M00023028A:A02
M00003839C:B05	M00004164D:D02	M00006777B:D10	M00023033A:E10
M00003845A:A05	M00004165C:E09	M00006779B:A11	M00023034C:E05
M00003846D:C12	M00004166A:F02	M00006779D:D03	M00023036D:C04
M00003857C:A03	M00004167C:F10	M00006780A:H12	M00023094A:C04
M00003858A:D01	M00004169A:B11	M00006789C:F04	M00023103A:E11
M00003860B:A07	M00004200B:B04	M00006790D:A05	M00006754B:D05

M00003868B:C07	M00004222A:H10	M00006796A:H10
M00003881D:D09	M00004223D:D07	M00006797B:D12
M00003883D:C03	M00004225D:F01	M00006801A:G05
M00003884B:E06	M00004228C:D11	M00006805A:E11
M00003886C:D10	M00004229C:G11	M00006805A:H09
M00003903C:A12	M00004239C:A07	M00006805B:C04
M00003912C:H01	M00004239C:C09	M00006807D:D08
M00003915B:G07	M00004240D:E06	M00006813A:C04
M00003920D:D09	M00004241B:B01	M00006822D:D05
M00003926B:E03	M00004243C:E10	M00006825C:D06
M00003934D:F01	M00004266A:F10	M00006831B:B04
M00003958C:C10	M00004266B:H06	M00006832A:F05
M00003965A:F07	M00004268C:F08	M00006832D:F10
M00003972C:F02	M00004268D:G07	M00006833B:E11
M00003974B:A04	M00004269A:B11	M00006872B:G01
M00003974C:A05	M00004269D:E08	M00006875D:D10
M00003975B:H09	M00004276C:E12	M00006879D:A10
M00003976C:C05	M00004277B:C06	M00006882D:F03
M00003980C:A11	M00004277C:H11	M00006884D:D06
M00003987A:C07	M00004279D:E02	M00006908C:A05
M00003988B:C10	M00004281B:B03	M00006921B:C02
M00003988C:A06	M00004284B:F07	M00006921B:E03
M00003989C:F01	M00004287B:B12	M00006949B:F03
M00004028C:D01	M00004287C:B06	M00006960A:G11
M00004029A:E01	M00004297D:B08	M00006966D:G03
M00004030A:E09	M00004332B:D02	M00006974B:D06
M00004031A:G05	M00004332B:E11	M00007013B:F02
M00004032D:D03	M00004346B:D06	M00007014D:C05
M00004033C:D10	M00004389C:E01	M00007014D:D04
M00004034A:E08	M00004403A:B05	M00007030A:G01
M00004035A:A10	M00004407D:B09	M00007030C:F08
M00004035B:H11	M00004419D:G01	M00007053B:C07

M00004035D:C05	M00004449D:H01	M00007065B:B12
M00004037B:A09	M00004463C:F11	M00007065D:C01
M00004037C:C05	M00004466A:E09	M00007075C:D08
M00004037D:B05	M00004469A:C12	M00007085A:B07
M00004044A:F08	M00004470C:A02	M00007118C:G02
M00004068A:F02	M00004498B:E01	M00007119B:H10
M00004068B:D04	M00004509A:H02	M00004824C:G09
M00004068D:B01	M00004605C:A09	M00004826A:E09
M00004069B:B01	M00004609C:C11	M00004839C:B01
M00004073D:E01	M00001378B:F06	M00004840C:F02
M00004075A:G10	M00005294C:G08	M00004840C:H05
M00004075C:C09	M00005294D:H02	M00004845D:E11
M00004076A:E02	M00005330C:F09	M00004846A:D02
M00004077D:D10	M00005333C:C08	M00004846D:H09
M00004078A:F03	M00005342B:G10	M00004854A:C09
M00004078C:A08	M00005352C:G09	M00004858D:E06
M00004084A:D11	M00005352D:E06	M00004999A:F01
M00004086A:A03	M00005353B:B09	M00004999B:D12
M00004086D:A07	M00005359B:G01	M00004999D:E01
M00004088A:F12	M00005359D:H08	M00005004B:C11
M00004089A:F02	M00005377A:A04	M00005005C:E06
M00004089A:G03	M00005377A:D05	M00005009B:A02
M00004093A:F03	M00005385C:D08	M00005015D:D11
M00004097C:A03	M00005388A:F07	M00005457D:C08
M00004102B:B04	M00005388D:B11	M00005519B:H04
M00004102C:F07	M00005392C:C04	M00005519C:F08
M00004103B:C09	M00005393A:E11	M00005531B:A03
M00004103C:F11	M00005394A:G07	M00005535B:F06
M00004104A:H09	M00005396B:C04	M00005587B:H02
M00004104D:C09	M00005399B:F02	M00005685A:A04
M00004108A:D04	M00005400A:D02	M00005706D:A09
M00004109B:A01	M00005403D:E11	M00005711A:H01

M00004126D:B11	M00005406D:B08	M00005798B:C11
M00004133C:B02	M00005411D:E05	M00005799C:C12
M00004182D:H03	M00005415D:G02	M00005805D:E06
M00004183A:D06	M00005417C:E10	M00005827B:H08
M00004186B:E05	M00005419A:D05	M00005828D:C09
M00004187C:H09	M00005419C:D09	M00005837A:D12
M00004188A:E05	M00005443D:C12	M00006751B:B11
M00004188A:E10	M00005447B:D02	M00006754B:D05
M00004190A:C12	M00005448D:E08	M00006756B:B08
M00004190C:G07	M00005450A:A02	M00006757D:E04
M00004190D:A10	M00005450A:B10	M00006758A:B12
M00004190D:G12	M00005450D:D02	M00006758D:C04
M00004198D:H04	M00005451A:E03	M00006834A:C08
M00004202B:F04	M00005456B:B07	M00006835B:F04
M00004202B:G09	M00005456B:E03	M00006837C:G06
M00004206C:G11	M00005460A:B10	M00006841D:A08
M00004213A:H12	M00005465C:H02	M00006855C:H02
M00004214A:D03	M00005466A:F12	M00006855D:H02
M00004218D:F12	M00005468B:D04	M00006859A:F06
M00004249C:E12	M00005470B:E01	M00006860B:H01
M00004249D:G02	M00005473D:E10	M00006886A:D06
M00004252D:A07	M00005483A:F05	M00006893C:B02
M00004253D:F09	M00005483D:A02	M00006893C:F02
M00004257C:A08	M00005487A:H01	M00006895D:E10
M00004262C:C01	M00005489A:F06	M00006917C:E07
M00001339B:E05	M00005493B:A12	M00006919B:C03
M00001341A:A11	M00005493B:E01	M00006923C:B01
M00001346A:B09	M00005497C:C10	M00006926A:H11
M00001346B:A07	M00005505A:C08	M00006934A:G02
M00001346B:G03	M00005508A:H01	M00006936B:E09
M00001346C:B07	M00005510B:D06	M00006936B:F10
M00001348A:G04	M00005528D:H06	M00006937B:F07

M00001348D:H08	M00005534A:G06	M00006937B:G09
M00001352C:E01	M00005539D:G07	M00006939B:E05
M00001362B:H09	M00005571A:E11	M00006953D:H11
M00001370A:B01	M00005619C:H10	M00006980A:F02
M00001370B:D04	M00005625D:C03	M00006986C:G11
M00001374C:C09	M00005626A:B11	M00006989B:C11
M00001376A:H02	M00005635B:A06	M00006990B:H09
M00001378B:F06	M00005635C:F11	M00006991A:E07
M00001380C:D10	M00005636C:D11	M00006991D:G07
M00001383C:C07	M00005637D:C05	M00006995C:A02
M00001384A:C09	M00005641B:E02	M00006997B:E06
M00001391D:A07	M00005645D:F08	M00006997D:B03
M00001391D:A09	M00005646C:B09	M00007006D:D04
M00001396C:G02	M00005646D:B03	M00007010B:C11
M00001397A:F10	M00005655D:C04	M00007010B:H03
M00001397B:E02	M00005703C:B01	M00007012B:D07
M00001397B:H11	M00005720B:D09	M00007031C:D01
M00001399D:F01	M00005722A:E09	M00007032A:F11
M00001400D:B08	M00005762D:A01	M00007033A:H05
M00001402C:E09	M00005783A:C05	M00007033D:F04
M00001406A:G12	M00005812C:F10	M00007036A:D02
M00001406D:B06	M00006581C:D02	M00007037B:D04
M00001408A:B02	M00006581D:H08	M00007084B:A05
M00001409C:D01	M00006582A:B09	M00007093A:F09
M00001411C:F02	M00006582D:E05	M00007099C:F09
M00001411D:C01	M00006592A:D03	M00007101A:A11
M00001412D:C03	M00006594D:F09	M00007107A:D11
M00001417B:C07	M00006596A:F07	M00007121C:H01
M00001417C:A09	M00006601D:F04	M00007129A:E04
M00001418A:C02	M00006604C:H10	M00007132B:B11
M00001421C:A03	M00006607B:E03	M00007134B:G07
M00001426A:C02	M00006607B:F04	M00007146D:G01

M00001427A:C05	M00006615D:F04	M00007148B:C06
M00001433A:F04	M00006616C:H09	M00007160C:B08
M00001434C:D05	M00006616D:C08	M00007161A:H03
M00001435C:H05	M00006617B:D09	M00007192C:H08
M00001438A:H10	M00006619B:C11	M00007200B:C02
M00001438B:H06		M00021619B:G10

Example 41: Source of Biological Materials and Overview of Novel Polynucleotides Expressed by the Biological Materials

cDNA libraries were constructed from either human colon cancer cell line Km12L4-A (Morikawa, et al., *Cancer Research* (1988) *48*:6863), KM12C (Morikawa et al. *Cancer Res.* (1988) *48*:1943-1948), or MDA-MB-231 (Brinkley et al. *Cancer Res.* (1980) *40*:3118-3129) was used to construct a cDNA library from mRNA isolated from the cells. Sequences expressed by these cell lines were isolated and analyzed; most sequences were about 275-300 nucleotides in length. The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B_2 surgical specimen (Morikawa *et al. Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman *et al. Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling *et al. Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as a model cell line for the study of colon cancer (see, *e.g.*, Moriakawa *et al.*, *supra*; Radinsky *et al. Clin. Cancer Res.* (1995) *1*:19; Yeatman *et al.*, (1995) *supra*; Yeatman *et al. Clin. Exp. Metastasis* (1996) *14*:246). The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma.

Example 42: Differential Expression of Polynucleotides of the Invention: Description of Libraries and Detection of Differential Expression

The relative expression levels of various polynucleotides isolated from the Example 41 were assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 64 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepared the cDNA library, the "nickname" of the library that is used in the tables below (in quotes), and the approximate number of clones in the library.

Table 64. Description of cDNA Libraries

Library (lib #)	Description	No. of Clones in Library
1	Human Colon Cell Line Km12 L4: High Metastatic Potential (derived from Km12C)	308731
2	Human Colon Cell Line Km12C: Low Metastatic Potential	284771

Library (lib #)	Description	No. of Clones in Library
3	Human Breast Cancer Cell Line MDA-MB-231: High Metastatic Potential; micro-metastases in lung	326937
4	Human Breast Cancer Cell Line MCF7: Non Metastatic	318979
8	Human Lung Cancer Cell Line MV-522: High Metastatic Potential	223620
9	Human Lung Cancer Cell Line UCP-3: Low Metastatic Potential	312503
12	Human microvascular endothelial cells (HMEC) - UNTREATED (PCR (OligodT) cDNA library)	41938
13	Human microvascular endothelial cells (HMEC) - bFGF TREATED (PCR (OligodT) cDNA library)	42100
14	Human microvascular endothelial cells (HMEC) - VEGF TREATED (PCR (OligodT) cDNA library)	42825
15	Normal Colon - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	282722
16	Colon Tumor - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	298831
17	Liver Metastasis from Colon Tumor of UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	303467
18	Normal Colon - UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	36216
19	Colon Tumor - UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	41388
20	Liver Metastasis from Colon Tumor of UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	30956
21	GRRpz Cells derived from normal prostate epithelium	164801
22	WOca Cells derived from Gleason Grade 4 prostate cancer epithelium	162088
23	Normal Lung Epithelium of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library)	306198
24	Primary tumor, Large Cell Carcinoma of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library)	309349

The KM12L4 and KM12C cell lines are described in Example 41 above. The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, *e.g.*, Chandrasekaran *et al.*, *Cancer Res.* (1979) *39*:870 (MDA-MB-231 and MCF-7); Gastpar *et al.*, *J Med Chem* (1998) *41*:4965 (MDA-MB-231 and MCF-7); Ranson *et al.*, *Br J Cancer* (1998) *77*:1586 (MDA-MB-231 and MCF-7); Kuang *et al.*, *Nucleic Acids Res* (1998) *26*:1116 (MDA-MB-231 and MCF-7); Varki *et al.*, *Int J Cancer* (1987) *40*:46 (UCP-3); Varki *et al.*, *Tumour Biol.* (1990) *11*:327; (MV-

522 and UCP-3); Varki *et al.*, *Anticancer Res.* (1990) *10*:637; (MV-522); Kelner *et al.*, *Anticancer Res* (1995) *15*:867 (MV-522); and Zhang *et al.*, *Anticancer Drugs* (1997) *8*:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-treated HMEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMEC were prepared by incubation with 20ng/ml VEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation. The GRRpz and WOca cell lines were provided by Dr. Donna M. Peehl, Department of Medicine, Stanford University School of Medicine. GRRpz was derived from normal prostate epithelium. The WOca cell line is a Gleason Grade 4 cell line.

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters. The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac *et al.*, *Genomics* (1996) *37*(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought together computationally. These groups of clones are termed "clusters". Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the "purity" of each cluster can be controlled. For example, artifacts of clustering may occur in computational clustering just as artifacts can occur in "wet-lab" screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1st), and the determining the number of cDNA clones corresponding to the selected cluster in

the second library (Clones in 2^{nd}). Differential expression of the selected cluster in the first library relative to the second library is expressed as a "ratio" of percent expression between the two libraries. In general, the "ratio" is calculated by: 1) calculating the percent expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the "number of clones" corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the "depth" of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5 , where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar, Biostatistical Analysis, Prentice Hall, Inc., USA, "Differences between Proportions," pp 296-298 (1974).

Using the methods and libraries described above, 37 of the isolated polynucleotides were identified as being differentially expressed across multiple libraries. Table 65 provides a list of these polynucleotides and their corresponding sequence names. The sequences of each of the above-referenced polynucleotides were determined using methods well known in the art. The sequences of the 37 polynucleotides, assigned SEQ ID NOS:8804-8840, are provided in the Sequence Listing below.

Table 65 Polynucleotides corresponding to differentially expressed genes

SEQ ID NO.	Sequence Name	SEQ ID NO.	Sequence Name
8804	13905	8823	RTA00000683F.1.19.1
8805	RTA00000281F.o.21.1	8824	RTA00000172A.d.9.3
8806	RTA00000348R.d.10.1	8825	RTA00000165A.d.16.1
8807	RTA00000177AF.d.22.3	8826	RTA00000188AR.d.05.1
8808	RTA00000684F.e.07.1	8827	RTA00000183AF.n.14.1
8809	RTA00000618F.p.24.1	8828	RTA00000346F.g.11.1
8810	RTA00000596F.d.12.1	8829	RTA00000183AR.n.14.1
8811	RTA00000421F.d.20.1	8830	RTA00000742F.g.08.1
8812	17090	8831	RTA00000689F.h.06.1
8813	RTA00000161A.1.7.1	8832	RTA00000185AF.b.9.1
8814	RTA00000155A.k.14.1	8833	RTA0000018SAF.b.9.2
8815	RTA00000163A.e.10.1	8834	RTA00000192AR.o.8.2
8816	RTA00000126A.o.15.2	8835	RTA00000192AF.o.8.1
8817	2546	8836	RTA00000685F.j.16.1
8818	RTA00000144A.p.8.1	8837	RTA00000621F.i.13.2
8819	RTA00000618F.k.16.1	8838	RTA00000685F.1.23.1
8820	RTA00000742F.o.19.1	8839	16405
8821	RTA00000148A.o.18.1	8840	028035A
8822	RTA00000619F.d.02.1		

The differential expression data for these sequences is provided below.

Example 43: Genes Differentially Expressed Genes in Non-Metastatic or Low Metastatic Potential Cancer Cells Versus High Metastatic Potential Cancer Cells

The relative levels of expression of genes corresponding to SEQ ID NO:8804-8840 across various libraries described in Table 64 are summarized in Table 66 below.

Table 66. Genes Differentially Expressed Across Multiple Library Comparisons

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	RATIO
8804	Low Met Breast (lib4) > High Met Breast (lib3)	5.38
8804	Low Met Colon (lib2) > High Met Colon (lib1)	6.14
8805	Low Met Colon (lib2) > High Met Colon (lib1)	3.56
8805	Low Met Breast (lib4) > High Met Breast (lib3)	2.73
8805	Normal Prostate (lib21) > Prostate Cancer (lib 22)	4.92
8806	Low Met Colon (lib2) > High Met Colon (lib1)	3.52
8806	Low Met Breast (lib4) > High Met Breast (lib3)	4.3
8807	Low Met Colon (lib2) > High Met Colon (lib1)	3.52
8807	Low Met Breast (lib4) > High Met Breast (lib3)	4.3
8808	High Met Lung (lib8) > Low Met Lung (lib9)	3.35
8808	Low Met Colon (lib2) > High Met Colon (lib1)	3.47
8808	Low Met Breast (lib4) > High Met Breast (lib3)	30.24

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	RATIO
8809	Low Met Breast (lib4) > High Met Breast (lib3)	30.24
8809	Low Met Colon (lib2) > High Met Colon (lib1)	3.47
8809	High Met Lung (lib8) > Low Met Lung (lib9)	3.35
8810	Low Met Colon (lib2) > High Met Colon (lib1)	3.47
8810	Low Met Breast (lib4) > High Met Breast (lib3)	30.24
8810	High Met Lung (lib8) > Low Met Lung (lib9)	3.35
8811	Low Met Breast (lib4) > High Met Breast (lib3)	2.42
8811	Low Met Colon (lib2) > High Met Colon (lib1)	2.63
8812	Low Met Colon (lib2) > High Met Colon (lib1)	2.49
8812	Low Met Breast (lib4) > High Met Breast (lib3)	2.19
8812	Low Met Lung (lib9) > High Met Lung (lib8)	3.07
8813	Low Met Breast (lib4) > High Met Breast (lib3)	41
8813	High Met Lung (lib8) > Low Met Lung (lib9)	2.29
8814	Low Met Breast (lib4) > High Met Breast (lib3)	7.35
8814	Normal Prostate (lib21) > Prostate Cancer (lib 22)	9.84
8815	High Met Breast (lib3) > Low Met Breast (lib4)	6.41
8815	High Met Colon (lib1) > Low Met Colon (lib2)	2.39
8816	High Met Colon (lib1) > Low Met Colon (lib2)	2.05
8816	High Met Breast (lib3) > Low Met Breast (lib4)	9.76
8817	Low Met Breast (lib4) > High Met Breast (lib3)	4.54
8817	High Met Lung (lib8) > Low Met Lung (lib9)	10.48
8817	Low Met Colon (lib2) > High Met Colon (lib1)	8.31
8818	Low Met Breast (lib4) > High Met Breast (lib3)	2.05
8818	Low Met Colon (lib2) > High Met Colon (lib1)	7.05
8819	Low Met Colon (lib2) > High Met Colon (lib1)	4.34
8819	Low Met Breast (lib4) > High Met Breast (lib3)	6.75
8820	Low Met Colon (lib2) > High Met Colon (lib1)	4.34
8820	Low Met Breast (lib4) > High Met Breast (lib3)	6.75
8821	Low Met Colon (lib2) > High Met Colon (lib1)	3.98
8821	Low Met Breast (lib4) > High Met Breast (lib3)	3.31
8821	Low Met Lung (lib9) > High Met Lung (lib8)	2.5
8822	Low Met Colon (lib2) > High Met Colon (lib1)	3.56
8822	Normal Prostate (lib21) > Prostate Cancer (lib 22)	4.92
8822	Low Met Breast (lib4) > High Met Breast (lib3)	2.73
8823	Normal Prostate (lib21) > Prostate Cancer (lib 22)	4.92
8823	Low Met Breast (lib4) > High Met Breast (lib3)	2.73
8823	Low Met Colon (lib2) > High Met Colon (lib1)	3.56
8824	Low Met Colon (lib2) > High Met Colon (lib1)	3.56
8824	Low Met Breast (lib4) > High Met Breast (lib3)	2.73
8824	Normal Prostate (lib21) > Prostate Cancer (lib 22)	4.92
8825	Low Met Colon (lib2) > High Met Colon (lib1)	3.52
8825	Low Met Breast (lib4) > High Met Breast (lib3)	3.55
8825	High Met Lung (lib8) > Low Met Lung (lib9)	17.7
8826	Low Met Colon (lib2) > High Met Colon (lib1)	3.25
8826	Low Met Breast (lib4) > High Met Breast (lib3)	3.07
8827	Low Met Breast (lib4) > High Met Breast (lib3)	3.07
8827	Low Met Colon (lib2) > High Met Colon (lib1)	3.25
8828	Low Met Colon (lib2) > High Met Colon (lib1)	3.25
8828	Low Met Breast (lib4) > High Met Breast (lib3)	3.07

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	RATIO
8829	Low Met Colon (lib2) > High Met Colon (lib1)	3.25
8829	Low Met Breast (lib4) > High Met Breast (lib3)	3.07
8830	Low Met Colon (lib2) > High Met Colon (lib1)	3.25
8830	Low Met Breast (lib4) > High Met Breast (lib3)	3.07
8831	Low Met Colon (lib2) > High Met Colon (lib1)	2.86
8831	Low Met Breast (lib4) > High Met Breast (lib3)	8.14
8832	Low Met Colon (lib2) > High Met Colon (lib1)	2.1
8832	Low Met Breast (lib4) > High Met Breast (lib3)	2.5
8833	Low Met Colon (lib2) > High Met Colon (lib1)	2.1
8833	Low Met Breast (lib4) > High Met Breast (lib3)	2.5
8834	Low Met Colon (lib2) > High Met Colon (lib1)	2.1
8834	Low Met Breast (lib4) > High Met Breast (lib3)	2.5
8835	Low Met Colon (lib2) > High Met Colon (lib1)	2.1
8835	Low Met Breast (lib4) > High Met Breast (lib3)	2.5
8836	Low Met Colon (lib2) > High Met Colon (lib1)	2.14
8836	Low Met Breast (lib4) > High Met Breast (lib3)	2.27
8837	Normal Prostate (lib21) > Prostate Cancer (lib 22)	5.9
8837	Low Met Colon (lib2) > High Met Colon (lib1)	2.1
8837	Low Met Breast (lib4) > High Met Breast (lib3)	2.18
8838	Normal Prostate (lib21) > Prostate Cancer (lib 22)	5.9
8838	Low Met Colon (lib2) > High Met Colon (lib1)	2.1
8838	Low Met Breast (lib4) > High Met Breast (lib3)	2.18
8839	Low Met Colon (lib2) > High Met Colon (lib1)	2.1
8839	Low Met Breast (lib4) > High Met Breast (lib3)	2.18
8839	Normal Prostate (lib21) > Prostate Cancer (lib 22)	5.9
8840	Low Met Colon (lib2) > High Met Colon (lib1)	2.17
8840	Low Met Breast (lib4) > High Met Breast (lib3)	2.9
8840	Low Met Lung (lib9) > High Met Lung (lib8)	3.4

Key for Table 66: High Met = high metastatic potential; Low Met = low metastatic potential; met = metastasized; tumor = non-metastasized tumor

The relative expression levels of the genes corresponding to the polynucleotides above can be exploited in diagnostic and prognostic assays. For example, where the polynucleotide corresponds to a gene that is expressed at a relatively higher level in a low metastatic potential cell relative to a high metastatic potential cell (or at a relatively higher level in normal cells or nonmetastasized tumor cells relatively to metastatic or high metastatic potential cancerous cells), expression of the gene can serve as a marker indicating low risk of metastasis and may encode a suppressor of metastasis. Where the polynucleotide corresponds to a gene expressed at a relatively higher level in a high metastatic potential cell relative to a low metastatic potential cell, expression of the gene can serve as a marker of metastatic potential, indicating the need for more aggressive therapy.

Example 44: Identification of a Gene and Protein Encoded by the Polynucleotide

SEQ ID NOS:8804-8840 were translated in all three reading frames, and the nucleotide sequences and translated amino acid sequences used as query sequences to search for homologous sequences in either the GenBank (nucleotide sequences) or Non-Redundant Protein (amino acid sequences) databases. Query and individual sequences were aligned using the BLAST 2.0 programs, available at the world wide web of the NCBI . (see also Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402). The sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity.

The results are provided in Table 67 below.

Table 67. Results of search of publicly available sequence databases using SEQ ID NOS:8804-8840 as query sequences

SEQ ID NO:	Description
8804	yt88d06.r1 Homo sapiens cDNA clone 231371 5'. (EST Accession No. H56522)
8805	za04c10.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone 291570 5' (EST Accession No. W03386)
8806	Homo sapiens heat shock factor binding protein 1 HSBP1 mRNA, complete cds (GenBank Accession No. AF068754)
8807	Homo sapiens heat shock factor binding protein 1 HSBP1 mRNA, complete cds (GenBank Accession No. AF068754)
8808	Homo sapiens CGI-122 protein mRNA, complete cds (GenBank Accession No. AF151880.1)
8809	Homo sapiens CGI-122 protein mRNA, complete cds (GenBank Accession No. AF151880.1)
8810	Homo sapiens CGI-122 protein mRNA, complete cds (GenBank Accession No. AF151880.1)
8811	zn42b05.s1 Stratagene endothelial cell 937223 Homo sapiens cDNA clone 550065 3' similar to SW:RPC9_YEAST P28000 DNA-DIRECTED RNA POLYMERASES I AND III 16 KD POLYPEPTIDE (EST Accession No. AA102570)
8812	yv31g09.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone 244384 5' similar to contains Alu repetitive element (EST Accession No. N72329)
8813	tz22h11.x1 NCI_CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2289381 3', mRNA sequence (EST Accession No. AI635233.1)
8814	zi02h12.r1 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA clone 429671 5' similar to contains Alu repetitive element (EST Accession No. AA011438)
8815	Human quiescin (Q6) mRNA
8816	Human Treacher Collins Syndrome
8817	Human mRNA for annexin IV (carbohydrate-binding protein p33/41)
8818	Human mRNA for TGIF protein
8819	Human MHC class I lymphocyte antigen (HLA-E) (HLA-6.2)
8820	Human HLA-E class I mRNA
8821	Human Mpv17 mRNA
8822	Human kidney cyclophilin C
8823	Human kidney cyclophilin C
8824	Human kidney cyclophilin C

SEQ ID NO:	Description
8825	Human mRNA for 26S proteasome subunit p55
8826	Human gamma-interferon-inducible protein (IP-30) mRNA
8827	Human gamma-interferon-inducible protein (IP-30) mRNA
8828	Human gamma-interferon-inducible protein (IP-30) mRNA
8829	Human gamma-interferon-inducible protein (IP-30) mRNA
8830	Human gamma-interferon-inducible protein (IP-30) mRNA
8831	Human Na+/H+ exchange regulatory co-factor (NHERF) mRNA
8832	Human mRNA for mitochondrial dodecenoyl-CoA delta-isomerase
8833	Human mRNA for mitochondrial dodecenoyl-CoA delta-isomerase
8834	Human mRNA for mitochondrial dodecenoyl-CoA delta-isomerase
8835	Human mRNA for mitochondrial dodecenoyl-CoA delta-isomerase
8836	Human (clone PSK-J3) cyclin-dependent protein kinase mRNA
8837	Human serine hydroxymethyltransferase mRNA
8838	Human serine hydroxymethyltransferase mRNA
8839	Human serine hydroxymethyltransferase mRNA
8840	Human DNA damage-inducible RNA binding protein (A18hnRNP).

Key: ES = EST database; GB = GenBank database

SEQ ID NO:8804 corresponds to a cDNA clone generated from an EST isolated from human pineal gland (Hillier *et al. Genome Res. (*1996) *6*(9):807-28).

SEQ ID NO:8805 corresponds to a sequence contained within a cDNA clone derived from an EST isolated from a human melanocyte 2NbHM.

SEQ ID NOS:8806 and 8807 correspond to a sequence encoding a human heat chock factor binding protein, HSBP-1, which acts as a negative regulator of the heat shock response through its interaction with heat shock factor 1 (HSF1) (Satyal *et al. Genes Dev.* (1998) *12*(13):1962-74). Briefly, HSF-1 responds to stress by undergoing conformational transition from an inert non-DNA binding monomer to an active trimed that exhibits rapid DNA binding and activity as a transcriptional activator. Attenuation of the inducible transcriptional response, which occurs during heat shock or upon recovery at non-stress conditions, involves dissociation of the HSF1 trimer and loss of activity. HSBP-1, a nuclear-localized, conserved, 76-amino-acid protein, contains two extended arrays of hydrophobic repeats that interact with HSF-1 heptad repeats of the active trimeric state of HSF1. During attenuation of HSF1 to the inert monomer, HSBP1 also associates with Hsp70. Through its interaction with HSF-1, HSBP1 negatively affects HSF-1 DNA-binding activity.

SEQ ID NOS:8808-8810 correspond to a gene encoding human CGI-122 protein.

SEQ ID NO:8811 corresponds to a cDNA clone generated from an EST isolated from human endothelial cells (Hillier *et al. Genome Res. (*1996) *6*(9):807-28).

SEQ ID NOS:8812 and 8814 correspond to a cDNA clone generated from an EST isolated from human fetal liver and spleen (Hillier *et al. Genome Res. (*1996) *6*(9):807-28).

SEQ ID NO:8813 corresponds to a sequence contained within a human cDNA clone isolated from moderately-differentiated endometrial adenocarcinoma.

The gene corresponding to SEQ ID NO:8816encodes human quiescin Q6 (Coppoch et al., 1998, Proc. Amer. Assoc. Can. Res. 39:471).

The gene corresponding to SEQ ID NO:8817 encodes a human Treacher Collins Syndrome protein. Treacher Collins Syndrome (TCS) is an autosomal dominant disorder of craniofacial development including hearing loss and cleft palate. The TCS gene (called Treacle) has been positionally cloned and has 26 exons exhibiting a low complexity serine/alanine-rich protein of about 144 kDa (Dixon et al., 1997, Genome Res. 7:223-234). Thirty-five mutations in the gene are reported from studies of individuals and families affected by Treacher Collins Syndrome (Edwards et al., 1997, Am. J. Human Genet. 60:515-524. Mutation in Treacle generally results in premature termination of the predicted protein (Nat. Genet. 12:130-136, 1996).

The gene corresponding to SEQ ID NO:8817 encodes human annexin IV (carbohydrate-binding protein p33/41). Annexins are a family of Ca2+ and phospholipid binding proteins. Annexin IV binds to glycosaminoglycans (GAGs) in a calcium-dependent manner (Kojima et al., 1996, J. Biol. Chem. 271:7679-7685; Ishitsuka et al., 1998, J. Biol. Chem. 273:9935-9941; and Satoh et al., 1997, Biol. Pharm. Bull. 20:224-229). Annexin IV is highly expressed in various human adenocarcinoma cell lines (Satoh et al., 1997, FEBS Lett. 405:107-110), and calcium-induced relocation of annexin IV is observed in a human osteosarcoma cell line (Mohiti et al., 1995, Mol. Membr. Biol. 12:321-329).

The gene corresponding to SEQ ID NO: 8818 encodes human TGIF protein (Bertolino et al., 1995, J. Biol. Chem. 270:31178-31188).

The gene corresponding to SEQ ID NO:8819 encodes human MHC Class I lymphocyte antigen (HLA-E) (HLA-6.2), as described by Koller et al., 1988, J. Immunol. 141:897-904.

The gene corresponding to SEQ ID NO:8820 encodes human HLA-E class I mRNA, as described by Mizuno et al., 1988, J. Immunol. 140:4024-4030.

The gene corresponding to SEQ ID NO:8821 is the human glomerulosclerosis gene Mpv17, as described by Karasawa, 1993, Hum. Mol. Genet. 11:1829-1834.

The gene corresponding to any one or more of SEQ ID NOS:8822-8824 encodes a human cyclophilin C (Schneider et al., 1994, Biochemistry 33:8218-8224).

The gene corresponding to SEQ ID NO:8825 encodes human 265 proteasome subunit p55. Human 26S proteasome is a heterodimer of p44.5 and p55 (Saito et al., 1997, Gene

203:241-250) and plays a major role in the non-lysosomal degradation of intracellular proteins (Mason et al., 1998, FEBS Lett. 430:269-274). Homologues of 26S proteasome subunits are regulators of transcription and translation as described in Aravind and Ponting, 1998, Protein Sci. 7:1250-1254. Proteasomes are cylindrical particles made up of a stack of four heptameric rings (Rivett et al., 1997, Mol. Biol. Rep. 24:99-102) and 26S proteasome has stringent organization of ATPases, as described in Seeger et al., 1997, Mol. Biol. Rep. 24:83-88. In mammalian cells, the proteasome is a site for degradation of proteins, as described in Goldberg et al., 1997, Biol. Chem. 378:131-140. In addition, proteolytic processing involving 26S proteasome occurs in lesions of Alzheimer's Disease and dementia with Lewy bodies (Fergusson et al., 1996, Neurosci. Lett. 219:167-170).

The gene corresponding to any one or more of SEQ ID NOS:8826-8830 encodes human gamma-interferon-inducible protein (IP-30), Luster et al., 1988, J. Biol. Chem. 263:12036-12043.

The gene corresponding to SEQ ID NO:8831 encodes human Na^+/H^+ exchange regulatory co-factor (NHEFR) (Murphy et al., 1998, J. Biol. Chem. in press).

The gene corresponding to any one or more of SEQ ID NOS:8832-8835 encodes human mitochondrial dodecenoyl-CoA delta-isomerase.

The gene corresponding to SEQ ID NO:8836 encodes human (clone PSK-J3) cyclin-dependent protein kinase (Hanks, 1987, Proc. Natl. Acad. Sci. 84:388-392).

The gene corresponding to any one or more of SEQ ID NOS:8837-8839 encodes human serine hydroxymethyltransferase. Human serine hydroxymethyltransferase is a pyridoxine enzyme that is low in resting lymphocytes but increases upon antigenic or mitogenic stimuli, such as in an immune response (Trakatellis et al., 1997, Postgrad. Med. J. 73:617-622, and Trakatellis et al., 1994, Postgrad. Med. J. 70(Suppl 1):S89-S92). The catalytic function of the protein is tested as described in Kim et al., 1997, Anal. Biochem. 253:201-209.

The polynucleotide comprising SEQ ID NO:8840 corresponds to a GenBank entry having accession number AF021336, an mRNA complete coding sequence for human DNA damage-inducible RNA binding protein (A18hnRNP). The p value of 1.9^{-113} indicates an extremely high level of similarity between the sequence of SEQ ID NO: 8840 and the identified GenBank sequence. Likewise, the protein search identified a high level of similarity (p value of 2.4^{-63}) between the amino acid translated from the second reading frame of the polynucleotide of SEQ ID NO: 8840 and the entry HUMCIRPA_1 for human mRNA for glycine-rich RNA binding protein cold-inducible RNA-binding protein (CIRP). The

search of DBEST identified accession number AA166551, murine CIRP, with a p value of 5.8^{-115}. CIRP is an 18kD protein induced in mouse cells by mild cold stress and consists of an N-terminal RNA-binding domain and a C-terminal glycine-rich domain (Nishiyama et al., 1997, J. Cell Biol. 137(4):899). Lowering the culture temperature of BALB/3T3 cells from 37°C to 32°C induces CIRP expression and impairs cell growth. Suppression of CIRP with antisense oligonucleotides alleviates the impaired growth, while overexpression of CIRP impairs growth at 37 °C and prolongs the G1 phase of the cell cycle (Nishiyama et al., *supra*). The cloning and characterization of human CIRP was described by Nishiyama et al., 1997, Gene 204(1-2):115).

Deposit Information. The materials described in Table 68 were deposited with the American Type Culture Collection (CMCC = Chiron Master Culture Collection).

Table 68. Cell Lines Deposited with ATCC

Cell Line	Deposit Date	ATCC Accession No.	CMCC Accession No.
KM12L4-A	March 19, 1998	CRL-12496	11606
Km12C	May 15, 1998	CRL-12533	11611
MDA-MB-231	May 15, 1998	CRL-12532	10583
MCF-7	October 9, 1998	CRL-12584	10377

The deposits described herein are provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby

Example 45: Source of Biological Materials and Overview of Novel Polynucleotides Expressed by the Biological Materials

cDNA libraries were constructed from either human colon cancer cell line Km12L4-A (Morikawa, et al., *Cancer Research* (1988) *48*:6863), KM12C (Morikawa et al. *Cancer Res.* (1988) *48*:1943-1948), or MDA-MB-231 (Brinkley et al. *Cancer Res.* (1980) *40*:3118-3129) was used to construct a cDNA library from mRNA isolated from the cells. Sequences expressed by these cell lines were isolated and analyzed; most sequences were about 275-300 nucleotides in length. The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B_2 surgical specimen (Morikawa *et al. Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman *et al. Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling *et al. Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as a model cell line for the study of colon cancer (see, *e.g.*, Moriakawa *et al.*, *supra*; Radinsky *et al. Clin. Cancer Res.* (1995) *1*:19; Yeatman *et al.*, (1995) *supra*; Yeatman *et al. Clin. Exp. Metastasis* (1996) *14*:246). The MDA-MB-231 cell line was originally isolated from pleural effusions (Cailleau, *J. Natl. Cancer. Inst.* (1974) *53*:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma.

The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol.* 266:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis Techniques" Adams *et al.*, eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie *et al. Comput. Chem.* (1993) 17:191). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats. Masking resulted in the elimination of 43 sequences. The remaining sequences were then used in a BLASTN vs. GenBank search; sequences that exhibited greater than 70% overlap, 99% identity, and a p value of less than 1×10^{-40} were discarded. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database

search: (1) unknown (no hits in the GenBank search), (2) weak similarity (greater than 45% identity and p value of less than 1 x 10^{-5}), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1 x 10^{-5}). Sequences having greater than 70% overlap, greater than 99% identity, and p value of less than 1 x 10^{-40} were discarded.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search was performed and sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1 x 10^{-40} were discarded. Sequences with a p value of less than 1 x 10^{-65} when compared to a database sequence of human origin were also excluded. Second, a BLASTN vs. Patent GeneSeq database was performed and sequences having greater than 99% identity, p value less than 1 x 10^{-40}, and greater than 99% overlap were discarded.

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1 x 10^{-111} in relation to a database sequence of human origin were specifically excluded. The final result provided the 982 sequences listed as SEQ ID NOS:8841-9785 in the accompanying Sequence Listing and summarized in Table 69A (inserted prior to claims). Each identified polynucleotide represents sequence from at least a partial mRNA transcript.

Table 69A provides: 1) the SEQ ID NO assigned to each sequence for use in the present specification; 2) the filing date of the U.S. priority application in which the sequence was first filed; 3) the attorney docket number assigned to the priority application (for internal use); 4) the SEQ ID NO assigned to the sequence in the priority application; 5) the sequence name used as an internal identifier of the sequence; and 6) the name assigned to the clone from which the sequence was isolated. Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOS: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene.

In order to confirm the sequences of SEQ ID NOS: 8841-9785, the clones were retrieved from a library using a robotic retrieval system, and the inserts of the retrieved clones re-sequenced. These "validation" sequences are provided as SEQ ID 9786:983-9799 in the Sequence Listing, and a summary of the "validation" sequences provided in Table 69B (inserted prior to claims). Table 69B provides: 1) the SEQ ID NO assigned to each sequence for use in the present specification; 2) the sample name assigned to the

"validation" sequence obtained; and 3) the name of the clone that contains the indicated "validation" sequence. "Validation" sequences can be correlated with the original sequences they validate by referring to Table 69A. Because the "validation" sequences are often longer than the original polynucleotide sequences and thus provide additional sequence information. All validation sequences can be obtained either from the corresponding clone or from a cDNA library described herein (*e.g.*, using primers designed from the sequence provided in the sequence listing).

Table 69A					
	Priority Appln Information				
SEQ ID NO:	Filed	Dkt No.	SEQ ID NO:	Sequence Name	Clone Name
8841	9/28/1998	1492.001	1	RTA00000617F.o.18.2	M00005513A:H01
8842	9/28/1998	1492.001	2	RTA00001075F.h.12.1	M00005434A:F11
8843	9/28/1998	1492.001	3	RTA00001076F.m.09.1	M00006946B:C08
8844	9/28/1998	1492.001	4	RTA00001075F.o.08.1	M00005628D:A10
8845	9/28/1998	1492.001	5	RTA00001064F.f.14.1	M00005465A:A07
8846	9/28/1998	1492.001	6	RTA00001075F.n.19.1	M00005614A:B07
8847	9/28/1998	1492.001	7	RTA00001075F.i.24.1	M00005453B:B06
8848	9/28/1998	1492.001	8	RTA00001075F.p.24.1	M00005721D:B03
8849	9/28/1998	1492.001	9	RTA00001075F.o.04.1	M00005621B:C09
8850	9/28/1998	1492.001	10	RTA00000616F.j.04.1	M00005412D:G07
8851	9/28/1998	1492.001	11	RTA00001064F.k.01.1	M00005708C:D11
8852	9/28/1998	1492.001	12	RTA00001064F.j.19.1	M00005657B:F11
8853	9/28/1998	1492.001	13	RTA00001065F.a.22.1	M00006920B:H07
8854	9/28/1998	1492.001	14	RTA00001076F.d.11.1	M00006623C:G07
8855	9/28/1998	1492.001	15	RTA00000615F.e.08.2	M00004872A:D07
8856	9/28/1998	1492.001	16	RTA00000617F.p.05.2	M00005515D:G02
8857	9/28/1998	1492.001	17	RTA00001076F.f.03.1	M00006668D:B10
8858	9/28/1998	1492.001	18	RTA00001064F.l.17.2	M00006582A:F12
8859	9/28/1998	1492.001	19	RTA00001076F.h.13.1	M00006745B:C05
8860	9/28/1998	1492.001	20	RTA00001075F.k.12.1	M00005482A:D08
8861	9/28/1998	1492.001	21	RTA00001076F.c.09.1	M00006594B:D05
8862	9/28/1998	1492.001	22	RTA00001076F.l.16.1	M00006919A:H12
8863	9/28/1998	1492.001	23	RTA00001076F.b.13.1	M00005825A:A10

8864	9/28/1998	1492.001	24	RTA00001065F.d.06.2	M00007078B:H04
8865	9/28/1998	1492.001	25	RTA00001075F.p.23.1	M00005721C:A12
8866	9/28/1998	1492.001	26	RTA00001075F.n.22.1	M00005616B:E11
8867	9/28/1998	1492.001	27	RTA00001075F.o.21.1	M00005648C:E10
8868	9/28/1998	1492.001	28	RTA00001065F.b.22.1	M00006968A:H05
8869	9/28/1998	1492.001	29	RTA00001075F.p.06.1	M00005698A:H12
8870	9/28/1998	1492.001	30	RTA00001076F.d.19.1	M00006630A:E05
8871	9/28/1998	1492.001	31	RTA00001075F.e.14.1	M00005375B:H03
8872	9/28/1998	1492.001	32	RTA00001065F.f.02.1	M00007186A:A12
8873	9/28/1998	1492.001	33	RTA00001064F.p.03.1	M00006814D:D09
8874	9/28/1998	1492.001	34	RTA00001076F.i.19.1	M00006813B:E04
8875	9/28/1998	1492.001	35	RTA00001077F.c.06.1	M00007157B:B04
8876	9/28/1998	1492.001	36	RTA00001064F.c.21.1	M00005366D:E12
8877	9/28/1998	1492.001	37	RTA00001065F.e.21.1	M00007177A:G07
8878	9/28/1998	1492.001	38	RTA00001076F.o.14.1	M00007038D:D01
8879	9/28/1998	1492.001	39	RTA00001064F.c.01.1	M00005327C:G08
8880	9/28/1998	1492.001	40	RTA00001064F.d.16.1	M00005397A:G08
8881	9/28/1998	1492.001	41	RTA00000615F.e.05.2	M00004870D:E05
8882	9/28/1998	1492.001	42	RTA00000616F.j.12.1	M00005413D:G12
8883	9/28/1998	1492.001	43	RTA00001075F.a.17.1	M00004852B:H08
8884	9/28/1998	1492.001	44	RTA00001076F.n.10.1	M00006989C:B01
8885	9/28/1998	1492.001	45	RTA00001075F.l.04.1	M00005505D:H08
8886	9/28/1998	1492.001	46	RTA00001075F.l.10.1	M00005509B:E10
8887	9/28/1998	1492.001	47	RTA00001075F.i.09.1	M00005444D:D01
8888	9/28/1998	1492.001	48	RTA00001075F.j.13.1	M00005464B:B08
8889	9/28/1998	1492.001	49	RTA00001076F.e.03.1	M00006635A:C01
8890	9/28/1998	1492.001	50	RTA00001076F.j.14.1	M00006837B:H12
8891	9/28/1998	1492.001	51	RTA00001075F.g.19.1	M00005418C:B09
8892	9/28/1998	1492.001	52	RTA00001075F.m.05.1	M00005538C:H11
8893	9/28/1998	1492.001	53	RTA00001076F.p.03.1	M00007046D:E10
8894	9/28/1998	1492.001	54	RTA00001075F.h.19.1	M00005435B:F01
8895	9/28/1998	1492.001	55	RTA00001075F.h.14.1	M00005434C:E02
8896	9/28/1998	1492.001	56	RTA00001076F.l.14.1	M00006917B:C05
8897	9/28/1998	1492.001	57	RTA00001075F.h.17.1	M00005434D:H02
8898	9/28/1998	1492.001	58	RTA00001075F.f.18.1	M00005396C:H04
8899	9/28/1998	1492.001	59	RTA00001076F.l.03.1	M00006894D:A07

8900	9/28/1998	1492.001	60	RTA00001065F.d.07.2	M00007079D:H01
8901	9/28/1998	1492.001	61	RTA00001075F.e.18.1	M00005377C:F07
8902	9/28/1998	1492.001	62	RTA00001065F.d.03.2	M00007065D:A03
8903	9/28/1998	1492.001	63	RTA00001076F.b.18.1	M00006577A:B01
8904	9/28/1998	1492.001	64	RTA00001075F.m.16.1	M00005569B:E04
8905	9/28/1998	1492.001	65	RTA00001076F.d.13.1	M00006627C:C02
8906	9/28/1998	1492.001	66	RTA00001076F.i.16.1	M00006805D:H12
8907	9/28/1998	1492.001	67	RTA00001076F.p.10.1	M00007064B:E09
8908	9/28/1998	1492.001	68	RTA00001064F.p.14.1	M00006835D:C08
8909	9/28/1998	1492.001	69	RTA00001077F.b.04.1	M00007126D:H01
8910	9/28/1998	1492.001	70	RTA00001076F.d.04.1	M00006619A:G11
8911	9/28/1998	1492.001	71	RTA00001077F.a.22.1	M00007121D:A11
8912	9/28/1998	1492.001	72	RTA00001077F.c.19.1	M00007178D:A10
8913	9/28/1998	1492.001	73	RTA00001065F.f.06.1	M00007197D:D12
8914	9/28/1998	1492.001	74	RTA00000616F.f.11.3	M00005395D:D11
8915	9/28/1998	1492.001	75	RTA00001064F.l.13.2	M00006577B:F01
8916	9/28/1998	1492.001	76	RTA00001064F.o.08.1	M00006757D:H04
8917	9/28/1998	1492.001	77	RTA00001075F.o.03.1	M00005621A:B05
8918	9/28/1998	1492.001	78	RTA00001064F.l.23.2	M00006596D:H02
8919	9/28/1998	1492.001	79	RTA00001076F.e.01.1	M00006631D:G09
8920	9/28/1998	1492.001	80	RTA00001075F.j.22.1	M00005473C:F02
8921	9/28/1998	1492.001	81	RTA00001076F.h.16.1	M00006757A:C09
8922	9/28/1998	1492.001	82	RTA00001075F.j.08.1	M00005459B:A01
8923	9/28/1998	1492.001	83	RTA00001064F.o.19.1	M00006795C:B12
8924	9/28/1998	1492.001	84	RTA00001064F.o.07.1	M00006756D:G07
8925	9/28/1998	1492.001	85	RTA00001076F.i.09.1	M00006790D:F10
8926	9/28/1998	1492.001	86	RTA00001076F.i.22.1	M00006815D:D11
8927	9/28/1998	1492.001	87	RTA00001076F.c.21.1	M00006613C:C02
8928	9/28/1998	1492.001	88	RTA00001076F.j.19.1	M00006846A:B03
8929	9/28/1998	1492.001	89	RTA00001064F.o.13.1	M00006779D:F03
8930	9/28/1998	1492.001	90	RTA00001077F.a.06.1	M00007101C:H01
8931	9/28/1998	1492.001	91	RTA00001064F.n.01.1	M00006664A:C05
8932	9/28/1998	1492.001	92	RTA00001064F.c.12.1	M00005358A:H03
8933	9/28/1998	1492.001	93	RTA00001077F.d.07.1	M00007196D:D02
8934	9/28/1998	1492.001	94	RTA00001077F.c.18.1	M00007177B:C02
8935	9/28/1998	1492.001	95	RTA00001064F.g.12.1	M00005490B:B02

8936	9/28/1998	1492.001	96	RTA00001075F.b.07.1	M00004866C:H08
8937	9/28/1998	1492.001	97	RTA00000617F.p.03.2	M00005515B:B08
8938	9/28/1998	1492.001	98	RTA00000616F.f.10.3	M00005395D:B12
8939	9/28/1998	1492.001	99	RTA00001064F.p.15.1	M00006840A:A12
8940	9/28/1998	1492.001	100	RTA00000617F.p.10.2	M00005516D:F12
8941	9/28/1998	1492.001	101	RTA00001076F.m.01.1	M00006925B:B02
8942	9/28/1998	1492.001	102	RTA00001075F.f.15.1	M00005395C:C11
8943	9/28/1998	1492.001	103	RTA00001075F.e.23.1	M00005385B:A10
8944	9/28/1998	1492.001	104	RTA00001076F.f.12.1	M00006688C:C12
8945	9/28/1998	1492.001	105	RTA00001075F.g.21.1	M00005420C:E03
8946	9/28/1998	1492.001	106	RTA00001076F.g.18.1	M00006727A:H12
8947	9/28/1998	1492.001	107	RTA00001075F.d.24.1	M00005363D:C05
8948	9/28/1998	1492.001	108	RTA00001075F.e.02.1	M00005364C:A02
8949	9/28/1998	1492.001	109	RTA00001075F.m.14.1	M00005563C:D05
8950	9/28/1998	1492.001	110	RTA00001064F.h.07.1	M00005520A:H11
8951	9/28/1998	1492.001	111	RTA00001065F.b.07.1	M00006936C:G11
8952	9/28/1998	1492.001	112	RTA00001065F.b.23.1	M00006968D:H02
8953	9/28/1998	1492.001	113	RTA00001064F.g.15.1	M00005497C:G08
8954	9/28/1998	1492.001	114	RTA00001064F.d.14.1	M00005390C:E05
8955	9/28/1998	1492.001	115	RTA00001064F.l.22.2	M00006595C:B08
8956	9/28/1998	1492.001	116	RTA00001064F.p.04.1	M00006816D:D08
8957	9/28/1998	1492.001	117	RTA00001076F.g.04.1	M00006712A:F01
8958	9/28/1998	1492.001	118	RTA00001075F.p.17.1	M00005709D:H05
8959	9/28/1998	1492.001	119	RTA00001075F.l.03.1	M00005505B:D10
8960	9/28/1998	1492.001	120	RTA00001076F.l.23.1	M00006925A:B09
8961	9/28/1998	1492.001	121	RTA00001076F.k.11.1	M00006874D:E01
8962	9/28/1998	1492.001	122	RTA00001076F.n.15.1	M00006994A:C12
8963	9/28/1998	1492.001	123	RTA00001075F.o.10.1	M00005629B:G06
8964	9/28/1998	1492.001	124	RTA00001075F.n.04.1	M00005589B:H12
8965	9/28/1998	1492.001	125	RTA00001075F.f.06.1	M00005388B:B02
8966	9/28/1998	1492.001	126	RTA00001076F.j.05.1	M00006823A:H06
8967	9/28/1998	1492.001	127	RTA00001076F.o.18.1	M00007041C:C05
8968	9/28/1998	1492.001	128	RTA00001064F.j.14.1	M00005648C:C11
8969	9/28/1998	1492.001	129	RTA00001064F.d.06.1	M00005376B:E08
8970	9/28/1998	1492.001	130	RTA00001077F.d.10.1	M00007200A:B12
8971	9/28/1998	1492.001	131	RTA00001065F.d.19.1	M00007109D:G01

8972	9/28/1998	1492.001	132	RTA00001064F.f.13.1	M00005464D:D07
8973	9/28/1998	1492.001	133	RTA00001075F.k.20.1	M00005493D:H12
8974	9/28/1998	1492.001	134	RTA00001075F.k.07.1	M00005479C:A05
8975	9/28/1998	1492.001	135	RTA00001075F.a.14.1	M00004847D:G01
8976	9/28/1998	1492.001	136	RTA00001076F.f.22.1	M00006704A:C11
8977	9/28/1998	1492.001	137	RTA00001076F.m.11.1	M00006949B:C07
8978	9/28/1998	1492.001	138	RTA00001064F.i.13.2	M00005618C:H11
8979	9/28/1998	1492.001	139	RTA00001076F.f.19.3	M00006694D:G06
8980	9/28/1998	1492.001	140	RTA00001076F.c.23.1	M00006617A:A06
8981	9/28/1998	1492.001	141	RTA00001077F.a.09.1	M00007107C:D02
8982	9/28/1998	1492.001	142	RTA00001064F.b.14.1	M00005020B:D10
8983	9/28/1998	1492.001	143	RTA00001075F.e.21.1	M00005382A:G09
8984	9/28/1998	1492.001	144	RTA00001075F.p.15.1	M00005705D:G09
8985	9/28/1998	1492.001	145	RTA00001076F.n.11.1	M00006991B:E05
8986	9/28/1998	1492.001	146	RTA00001065F.e.18.1	M00007161C:D12
8987	9/28/1998	1492.001	147	RTA00000615F.e.06.2	M00004871C:C04
8988	9/28/1998	1492.001	148	RTA00001064F.a.04.2	M00004821D:C03
8989	9/28/1998	1492.001	149	RTA00001075F.j.18.1	M00005469A:D10
8990	9/28/1998	1492.001	150	RTA00001077F.c.05.1	M00007156D:E11
8991	9/28/1998	1492.001	151	RTA00001075F.g.22.1	M00005420C:E10
8992	9/28/1998	1492.001	152	RTA00001077F.a.08.1	M00007104D:D10
8993	9/28/1998	1492.001	153	RTA00001077F.c.15.1	M00007172D:H03
8994	9/28/1998	1492.001	154	RTA00001077F.c.16.1	M00007175B:B11
8995	9/28/1998	1492.001	155	RTA00001077F.b.15.1	M00007141A:G08
8996	9/28/1998	1492.001	156	RTA00001077F.c.17.1	M00007175D:G02
8997	9/28/1998	1492.001	157	RTA00001077F.a.14.1	M00007116A:C08
8998	9/28/1998	1492.001	158	RTA00001075F.i.02.1	M00005438D:A08
8999	9/28/1998	1492.001	159	RTA00001075F.l.11.1	M00005509D:G05
9000	9/28/1998	1492.001	160	RTA00001064F.d.20.1	M00005403A:D12
9001	9/28/1998	1492.001	161	RTA00001076F.h.10.1	M00006740A:A06
9002	9/28/1998	1492.001	162	RTA00001075F.k.21.1	M00005494C:F08
9003	9/28/1998	1492.001	163	RTA00001075F.i.21.1	M00005450C:G09
9004	9/28/1998	1492.001	164	RTA00001076F.p.24.1	M00007093C:C11
9005	9/28/1998	1492.001	165	RTA00001075F.f.03.1	M00005385D:B08
9006	9/28/1998	1492.001	166	RTA00001065F.d.18.2	M00007107A:H08
9007	9/28/1998	1492.001	167	RTA00001076F.o.05.1	M00007026A:A03

9008	9/28/1998	1492.001	168	RTA00001075F.d.10.1	M00005353C:H01
9009	9/28/1998	1492.001	169	RTA00001064F.d.07.1	M00005378B:B04
9010	9/28/1998	1492.001	170	RTA00001065F.b.11.1	M00006945D:A07
9011	9/28/1998	1492.001	171	RTA00001076F.g.17.1	M00006726D:H10
9012	9/28/1998	1492.001	172	RTA00001065F.a.21.1	M00006918D:G08
9013	9/28/1998	1492.001	173	RTA00001077F.d.12.1	M00007203C:E06
9014	9/28/1998	1492.001	174	RTA00001064F.g.08.1	M00005481C:H05
9015	9/28/1998	1492.001	175	RTA00001064F.f.02.1	M00005449D:D04
9016	9/28/1998	1492.001	176	RTA00001075F.a.02.1	M00004825A:G12
9017	9/28/1998	1492.001	177	RTA00001064F.b.16.1	M00005296B:H07
9018	9/28/1998	1492.001	178	RTA00001077F.c.02.1	M00007152A:A10
9019	9/28/1998	1492.001	179	RTA00001064F.g.04.1	M00005480C:A04
9020	9/28/1998	1492.001	180	RTA00001075F.c.12.1	M00005305A:H01
9021	9/28/1998	1492.001	181	RTA00001064F.o.04.1	M00006752C:D04
9022	9/28/1998	1492.001	182	RTA00001077F.a.21.1	M00007121A:G04
9023	9/28/1998	1492.001	183	RTA00001075F.f.11.1	M00005392C:B03
9024	9/28/1998	1492.001	184	RTA00001064F.k.24.2	M00005820A:H11
9025	9/28/1998	1492.001	185	RTA00001075F.d.02.1	M00005342D:E04
9026	9/28/1998	1492.001	186	RTA00001076F.c.13.1	M00006600D:G07
9027	9/28/1998	1492.001	187	RTA00001075F.b.15.1	M00004872C:G03
9028	9/28/1998	1492.001	188	RTA00001064F.f.09.1	M00005461C:D11
9029	9/28/1998	1492.001	189	RTA00001075F.g.14.1	M00005416B:A01
9030	9/28/1998	1492.001	190	RTA00001075F.f.17.1	M00005396A:C01
9031	9/28/1998	1492.001	191	RTA00001076F.l.05.1	M00006895D:A02
9032	9/28/1998	1492.001	192	RTA00001076F.o.02.1	M00007019B:G01
9033	9/28/1998	1492.001	193	RTA00001064F.b.07.1	M00005000A:H05
9034	9/28/1998	1492.001	194	RTA00001075F.d.17.1	M00005358B:D10
9035	9/28/1998	1492.001	195	RTA00000624F.f.12.2	M00005607A:C08
9036	9/28/1998	1492.001	196	RTA00001075F.c.22.1	M00005342B:G01
9037	9/28/1998	1492.001	197	RTA00001065F.a.17.1	M00006914C:D07
9038	9/28/1998	1492.001	198	RTA00001075F.b.02.1	M00004859D:D01
9039	9/28/1998	1492.001	199	RTA00001077F.c.12.1	M00007167C:B10
9040	9/28/1998	1492.001	200	RTA00001077F.c.20.1	M00007179B:H04
9041	9/28/1998	1492.001	201	RTA00001076F.m.04.1	M00006934B:B11
9042	9/28/1998	1492.001	202	RTA00001076F.j.22.1	M00006859D:E11
9043	9/28/1998	1492.001	203	RTA00001076F.k.13.1	M00006882C:D03

9044	9/28/1998	1492.001	204	RTA00001075F.k.14.1	M00005485C:F09
9045	9/28/1998	1492.001	205	RTA00001076F.f.10.1	M00006680D:A01
9046	9/28/1998	1492.001	206	RTA00001064F.o.05.1	M00006755C:C03
9047	9/28/1998	1492.001	207	RTA00001064F.l.05.2	M00005826B:F10
9048	9/28/1998	1492.001	208	RTA00001076F.p.04.1	M00007047D:C02
9049	9/28/1998	1492.001	209	RTA00001064F.l.04.1	M00005822D:C05
9050	9/28/1998	1492.001	210	RTA00001076F.c.03.1	M00006584D:D01
9051	9/28/1998	1492.001	211	RTA00001064F.m.06.1	M00006621B:B06
9052	9/28/1998	1492.001	212	RTA00001075F.k.15.1	M00005486A:F07
9053	9/28/1998	1492.001	213	RTA00001064F.d.08.1	M00005378C:B12
9054	9/28/1998	1492.001	214	RTA00001077F.d.11.1	M00007202A:A09
9055	9/28/1998	1492.001	215	RTA00001077F.b.14.1	M00007140C:G12
9056	9/28/1998	1492.001	216	RTA00001075F.k.04.1	M00005476D:A11
9057	9/28/1998	1492.001	217	RTA00001064F.n.03.1	M00006678C:B07
9058	9/28/1998	1492.001	218	RTA00001075F.i.12.1	M00005446B:D10
9059	9/28/1998	1492.001	219	RTA00001075F.f.04.1	M00005386C:G01
9060	9/28/1998	1492.001	220	RTA00001076F.n.14.1	M00006993B:F02
9061	9/28/1998	1492.001	221	RTA00001064F.k.19.2	M00005810B:C07
9062	9/28/1998	1492.001	222	RTA00001076F.d.20.1	M00006630A:E09
9063	9/28/1998	1492.001	223	RTA00001077F.b.20.1	M00007145C:B05
9064	9/28/1998	1492.001	224	RTA00001076F.f.11.1	M00006688A:F09
9065	9/28/1998	1492.001	225	RTA00001065F.d.01.1	M00007047C:H04
9066	9/28/1998	1492.001	226	RTA00001075F.g.12.1	M00005413B:B02
9067	9/28/1998	1492.001	227	RTA00001064F.a.09.2	M00004841C:H03
9068	9/28/1998	1492.001	228	RTA00001064F.k.20.2	M00005810B:G02
9069	9/28/1998	1492.001	229	RTA00001064F.b.17.1	M00005296D:G03
9070	9/28/1998	1493.001	1	RTA00001073F.f.17.1	M00004087A:H06
9071	9/28/1998	1493.001	2	RTA00001073F.l.02.1	M00004168D:F05
9072	9/28/1998	1493.001	3	RTA00001072F.i.07.3	M00003845B:A04
9073	9/28/1998	1493.001	4	RTA00001071F.i.23.3	M00001477A:G02
9074	9/28/1998	1493.001	5	RTA00000611F.e.04.2	M00004170C:H06
9075	9/28/1998	1493.001	6	RTA00001062F.f.19.1	M00003888C:G08
9076	9/28/1998	1493.001	7	RTA00001073F.l.22.1	M00004176B:H09
9077	9/28/1998	1493.001	8	RTA00001063F.l.10.1	M00004410A:F06
9078	9/28/1998	1493.001	9	RTA00001062F.l.13.1	M00004034A:A05
9079	9/28/1998	1493.001	10	RTA00001074F.l.10.1	M00004495D:A05

9080	9/28/1998	1493.001	11	RTA00001061F.d.01.1	M00001389C:E01
9081	9/28/1998	1493.001	12	RTA00001072F.j.04.2	M00003861D:G10
9082	9/28/1998	1493.001	13	RTA00001073F.d.04.1	M00004048C:C02
9083	9/28/1998	1493.001	14	RTA00001061F.j.09.1	M00001507A:H06
9084	9/28/1998	1493.001	15	RTA00001071F.h.16.1	M00001450D:H12
9085	9/28/1998	1493.001	16	RTA00001062F.o.17.1	M00004108B:D04
9086	9/28/1998	1493.001	17	RTA00001073F.c.20.1	M00004046C:A04
9087	9/28/1998	1493.001	18	RTA00001063F.k.14.1	M00004381A:E10
9088	9/28/1998	1493.001	19	RTA00000611F.e.18.2	M00004171D:H10
9089	9/28/1998	1493.001	20	RTA00001072F.a.18.2	M00001655C:F07
9090	9/28/1998	1493.001	21	RTA00001072F.b.04.2	M00001660A:B10
9091	9/28/1998	1493.001	22	RTA00001074F.g.19.1	M00004372A:A08
9092	9/28/1998	1493.001	23	RTA00001072F.i.09.3	M00003845C:F08
9093	9/28/1998	1493.001	24	RTA00001072F.a.21.2	M00001657D:D07
9094	9/28/1998	1493.001	25	RTA00001072F.m.18.3	M00003916D:A10
9095	9/28/1998	1493.001	26	RTA00001061F.b.04.1	M00001360B:F09
9096	9/28/1998	1493.001	27	RTA00001072F.o.06.2	M00003935A:C04
9097	9/28/1998	1493.001	28	RTA00001072F.n.19.3	M00003931A:G01
9098	9/28/1998	1493.001	29	RTA00001073F.e.08.1	M00004068A:A03
9099	9/28/1998	1493.001	30	RTA00001074F.g.22.1	M00004373D:G10
9100	9/28/1998	1493.001	31	RTA00001073F.c.01.1	M00004030C:E05
9101	9/28/1998	1493.001	32	RTA00001074F.f.15.1	M00004360B:B08
9102	9/28/1998	1493.001	33	RTA00001074F.f.01.1	M00004350A:C04
9103	9/28/1998	1493.001	34	RTA00001074F.d.08.1	M00004318D:D07
9104	9/28/1998	1493.001	35	RTA00001072F.f.11.2	M00003788D:E06
9105	9/28/1998	1493.001	36	RTA00001074F.e.05.1	M00004337A:A07
9106	9/28/1998	1493.001	37	RTA00001072F.g.05.2	M00003803B:G12
9107	9/28/1998	1493.001	38	RTA00001071F.j.04.3	M00001479D:B10
9108	9/28/1998	1493.001	39	RTA00001074F.j.05.1	M00004415A:A01
9109	9/28/1998	1493.001	40	RTA00001074F.j.04.1	M00004414D:C11
9110	9/28/1998	1493.001	41	RTA00001073F.e.06.1	M00004067C:C10
9111	9/28/1998	1493.001	42	RTA00001071F.d.14.1	M00001389A:F03
9112	9/28/1998	1493.001	43	RTA00001071F.f.12.1	M00001418C:F06
9113	9/28/1998	1493.001	44	RTA00001061F.m.13.1	M00001601D:A03
9114	9/28/1998	1493.001	45	RTA00001061F.e.17.1	M00001418A:A02
9115	9/28/1998	1493.001	46	RTA00001071F.m.09.3	M00001563A:F04

9116	9/28/1998	1493.001	47	RTA00001062F.l.05.1	M00004029D:H03
9117	9/28/1998	1493.001	48	RTA00001073F.i.02.2	M00004125B:A02
9118	9/28/1998	1493.001	49	RTA00001063F.l.04.1	M00004404C:B03
9119	9/28/1998	1493.001	50	RTA00001063F.l.14.1	M00004412A:G05
9120	9/28/1998	1493.001	51	RTA00001063F.e.05.1	M00004232D:G11
9121	9/28/1998	1493.001	52	RTA00001062F.f.06.1	M00003880A:G10
9122	9/28/1998	1493.001	53	RTA00001072F.b.23.2	M00001683B:F12
9123	9/28/1998	1493.001	54	RTA00001073F.a.13.1	M00003989D:A02
9124	9/28/1998	1493.001	55	RTA00001074F.h.16.1	M00004386C:C03
9125	9/28/1998	1493.001	56	RTA00001073F.a.15.1	M00003991A:D05
9126	9/28/1998	1493.001	57	RTA00001073F.k.01.1	M00004152A:F03
9127	9/28/1998	1493.001	58	RTA00001072F.l.19.2	M00003901B:C02
9128	9/28/1998	1493.001	59	RTA00001072F.i.15.3	M00003848A:E08
9129	9/28/1998	1493.001	60	RTA00001072F.i.05.3	M00003844D:B02
9130	9/28/1998	1493.001	61	RTA00001074F.m.06.1	M00004603D:D09
9131	9/28/1998	1493.001	62	RTA00001062F.m.15.1	M00004063B:B12
9132	9/28/1998	1493.001	63	RTA00001074F.d.19.1	M00004326D:D06
9133	9/28/1998	1493.001	64	RTA00001073F.j.02.1	M00004140B:C02
9134	9/28/1998	1493.001	65	RTA00001071F.l.11.1	M00001545D:F12
9135	9/28/1998	1493.001	66	RTA00001074F.f.12.1	M00004356C:D02
9136	9/28/1998	1493.001	67	RTA00001073F.h.03.1	M00004110A:G03
9137	9/28/1998	1493.001	68	RTA00001074F.a.19.1	M00004275A:H07
9138	9/28/1998	1493.001	69	RTA00001063F.g.15.1	M00004292A:C08
9139	9/28/1998	1493.001	70	RTA00001061F.a.09.1	M00001345C:B10
9140	9/28/1998	1493.001	71	RTA00001063F.f.23.1	M00004284A:C09
9141	9/28/1998	1493.001	72	RTA00001073F.e.10.1	M00004069A:E04
9142	9/28/1998	1493.001	73	RTA00001073F.g.15.1	M00004103A:E06
9143	9/28/1998	1493.001	74	RTA00001073F.n.20.1	M00004209B:G01
9144	9/28/1998	1493.001	75	RTA00001073F.g.11.1	M00004099C:F04
9145	9/28/1998	1493.001	76	RTA00001071F.p.05.1	M00001630A:E08
9146	9/28/1998	1493.001	77	RTA00001073F.l.19.1	M00004175D:D05
9147	9/28/1998	1493.001	78	RTA00001074F.j.17.1	M00004426B:H06
9148	9/28/1998	1493.001	79	RTA00001074F.b.22.1	M00004292A:F03
9149	9/28/1998	1493.001	80	RTA00001071F.d.19.1	M00001391C:B05
9150	9/28/1998	1493.001	81	RTA00001062F.j.02.1	M00003960D:E09
9151	9/28/1998	1493.001	82	RTA00001072F.b.09.2	M00001664D:E02

9152	9/28/1998	1493.001	83	RTA00001073F.b.08.1	M00003998C:D04
9153	9/28/1998	1493.001	84	RTA00001062F.j.19.1	M00003977D:H04
9154	9/28/1998	1493.001	85	RTA00001062F.m.18.1	M00004066D:C02
9155	9/28/1998	1493.001	86	RTA00001062F.b.02.1	M00003775C:C01
9156	9/28/1998	1493.001	87	RTA00001061F.d.20.1	M00001401B:A02
9157	9/28/1998	1493.001	88	RTA00001071F.n.05.3	M00001579C:E07
9158	9/28/1998	1493.001	89	RTA00001073F.l.04.1	M00004170B:G04
9159	9/28/1998	1493.001	90	RTA00001071F.h.04.1	M00001442D:D09
9160	9/28/1998	1493.001	91	RTA00001062F.o.11.1	M00004104C:F06
9161	9/28/1998	1493.001	92	RTA00001062F.i.10.1	M00003939B:C02
9162	9/28/1998	1493.001	93	RTA00001071F.g.16.1	M00001431A:F03
9163	9/28/1998	1493.001	94	RTA00001061F.d.06.1	M00001392A:F02
9164	9/28/1998	1493.001	95	RTA00001071F.m.01.3	M00001561A:G10
9165	9/28/1998	1493.001	96	RTA00001062F.n.06.1	M00004081A:E11
9166	9/28/1998	1493.001	97	RTA00001061F.d.14.1	M00001397D:G04
9167	9/28/1998	1493.001	98	RTA00001061F.j.10.1	M00001507D:F09
9168	9/28/1998	1493.001	99	RTA00001063F.c.07.1	M00004185B:H03
9169	9/28/1998	1493.001	100	RTA00001061F.j.12.1	M00001513B:F05
9170	9/28/1998	1493.001	101	RTA00001061F.o.22.1	M00001678A:B10
9171	9/28/1998	1493.001	102	RTA00001071F.e.03.1	M00001395D:B04
9172	9/28/1998	1493.001	103	RTA00001072F.e.13.2	M00003772C:F12
9173	9/28/1998	1493.001	104	RTA00001062F.i.03.1	M00003928D:A04
9174	9/28/1998	1493.001	105	RTA00001072F.d.20.2	M00003761C:C05
9175	9/28/1998	1493.001	106	RTA00001074F.g.16.1	M00004371B:A05
9176	9/28/1998	1493.001	107	RTA00001074F.f.09.1	M00004353D:C06
9177	9/28/1998	1493.001	108	RTA00001071F.k.12.1	M00001505C:C10
9178	9/28/1998	1493.001	109	RTA00001074F.f.13.1	M00004357A:B10
9179	9/28/1998	1493.001	110	RTA00001071F.e.08.1	M00001397C:F01
9180	9/28/1998	1493.001	111	RTA00001073F.h.11.1	M00004117D:F06
9181	9/28/1998	1493.001	112	RTA00001072F.o.14.2	M00003937D:F09
9182	9/28/1998	1493.001	113	RTA00001074F.c.11.1	M00004298A:H09
9183	9/28/1998	1493.001	114	RTA00001074F.g.08.1	M00004368A:G11
9184	9/28/1998	1493.001	115	RTA00001073F.a.18.1	M00003993C:G11
9185	9/28/1998	1493.001	116	RTA00001073F.f.19.1	M00004090A:B11
9186	9/28/1998	1493.001	117	RTA00001072F.l.20.2	M00003902C:D02
9187	9/28/1998	1493.001	118	RTA00001073F.b.06.1	M00003997D:G03

9188	9/28/1998	1493.001	119	RTA00001062F.o.14.1	M00004105C:C05
9189	9/28/1998	1493.001	120	RTA00001071F.i.04.3	M00001457D:E08
9190	9/28/1998	1493.001	121	RTA00001074F.a.23.1	M00004278C:H11
9191	9/28/1998	1493.001	122	RTA00001073F.c.04.1	M00004034A:G03
9192	9/28/1998	1493.001	123	RTA00001072F.h.18.2	M00003833D:F11
9193	9/28/1998	1493.001	124	RTA00001074F.i.06.1	M00004403A:A02
9194	9/28/1998	1493.001	125	RTA00001063F.e.09.1	M00004240A:D03
9195	9/28/1998	1493.001	126	RTA00001061F.d.03.1	M00001390C:H05
9196	9/28/1998	1493.001	127	RTA00001063F.d.23.1	M00004225A:E03
9197	9/28/1998	1493.001	128	RTA00001063F.k.08.1	M00004378A:H10
9198	9/28/1998	1493.001	129	RTA00001062F.b.04.1	M00003776B:F08
9199	9/28/1998	1493.001	130	RTA00001063F.b.18.1	M00004178B:F07
9200	9/28/1998	1493.001	131	RTA00001062F.b.11.1	M00003788B:C08
9201	9/28/1998	1493.001	132	RTA00001074F.l.23.1	M00004504C:G07
9202	9/28/1998	1493.001	133	RTA00001063F.m.08.1	M00004444C:H11
9203	9/28/1998	1493.001	134	RTA00001071F.l.13.2	M00001549C:F10
9204	9/28/1998	1493.001	135	RTA00001072F.p.19.2	M00003973A:D09
9205	9/28/1998	1493.001	136	RTA00001071F.k.17.1	M00001517C:A10
9206	9/28/1998	1493.001	137	RTA00001072F.o.24.2	M00003943B:C12
9207	9/28/1998	1493.001	138	RTA00001074F.a.20.1	M00004276A:C06
9208	9/28/1998	1493.001	139	RTA00001073F.c.16.1	M00004043C:A06
9209	9/28/1998	1493.001	140	RTA00001074F.j.10.1	M00004422C:A01
9210	9/28/1998	1493.001	141	RTA00001063F.n.16.1	M00004498D:F02
9211	9/28/1998	1493.001	142	RTA00001071F.o.16.1	M00001615A:D01
9212	9/28/1998	1493.001	143	RTA00001073F.k.16.1	M00004165D:H12
9213	9/28/1998	1493.001	144	RTA00001062F.e.14.1	M00003856A:H10
9214	9/28/1998	1493.001	145	RTA00001071F.h.22.1	M00001454D:H09
9215	9/28/1998	1493.001	146	RTA00001071F.o.18.1	M00001618C:E01
9216	9/28/1998	1493.001	147	RTA00001062F.p.19.1	M00004140D:E03
9217	9/28/1998	1493.001	148	RTA00001062F.d.04.1	M00003818C:D02
9218	9/28/1998	1493.001	149	RTA00001072F.n.22.3	M00003933A:B04
9219	9/28/1998	1493.001	150	RTA00001063F.c.11.1	M00004187A:B05
9220	9/28/1998	1493.001	151	RTA00001061F.j.22.1	M00001531B:A03
9221	9/28/1998	1493.001	152	RTA00001062F.d.08.1	M00003820C:E08
9222	9/28/1998	1493.001	153	RTA00001062F.f.02.1	M00003877C:G01
9223	9/28/1998	1493.001	154	RTA00001062F.d.24.1	M00003839D:C03

9224	9/28/1998	1493.001	155	RTA00001074F.h.24.1	M00004391C:F12
9225	9/28/1998	1493.001	156	RTA00001071F.a.10.1	M00001341A:H10
9226	9/28/1998	1493.001	157	RTA00001074F.k.13.1	M00004449B:B05
9227	9/28/1998	1493.001	158	RTA00001072F.k.16.2	M00003884C:G09
9228	9/28/1998	1493.001	159	RTA00001073F.k.09.1	M00004158C:B01
9229	9/28/1998	1493.001	160	RTA00001074F.b.14.1	M00004288D:E07
9230	9/28/1998	1493.001	161	RTA00001073F.k.08.1	M00004157C:E06
9231	9/28/1998	1493.001	162	RTA00001074F.i.17.1	M00004406D:E11
9232	9/28/1998	1493.001	163	RTA00001074F.k.10.1	M00004447A:A10
9233	9/28/1998	1493.001	164	RTA00001062F.p.14.1	M00004135D:D01
9234	9/28/1998	1493.001	165	RTA00001071F.m.15.3	M00001569A:H01
9235	9/28/1998	1493.001	166	RTA00001074F.h.15.1	M00004385D:D06
9236	9/28/1998	1493.001	167	RTA00001062F.i.09.1	M00003935D:E04
9237	9/28/1998	1493.001	168	RTA00000611F.e.06.2	M00004170D:C06
9238	9/28/1998	1493.001	169	RTA00001062F.d.19.1	M00003835B:C05
9239	9/28/1998	1493.001	170	RTA00001062F.o.15.1	M00004107A:E02
9240	9/28/1998	1493.001	171	RTA00001071F.a.07.1	M00001340C:A08
9241	9/28/1998	1493.001	172	RTA00001062F.d.07.1	M00003820B:G04
9242	9/28/1998	1493.001	173	RTA00001074F.j.11.1	M00004423A:B05
9243	9/28/1998	1493.001	174	RTA00001071F.m.11.3	M00001565C:F06
9244	9/28/1998	1493.001	175	RTA00001062F.i.01.1	M00003926A:D01
9245	9/28/1998	1493.001	176	RTA00001072F.g.08.2	M00003804D:F12
9246	9/28/1998	1493.001	177	RTA00001071F.n.16.1	M00001594A:H01
9247	9/28/1998	1493.001	178	RTA00001062F.a.09.1	M00003756D:B09
9248	9/28/1998	1493.001	179	RTA00001073F.h.08.1	M00004114C:B09
9249	9/28/1998	1493.001	180	RTA00001073F.e.03.1	M00004064B:G03
9250	9/28/1998	1493.001	181	RTA00001073F.c.23.1	M00004048A:E10
9251	9/28/1998	1493.001	182	RTA00001074F.l.15.1	M00004498D:A11
9252	9/28/1998	1493.001	183	RTA00001073F.l.21.1	M00004176A:H05
9253	9/28/1998	1493.001	184	RTA00001071F.d.15.1	M00001389B:B12
9254	9/28/1998	1493.001	185	RTA00001073F.i.08.1	M00004127C:C08
9255	9/28/1998	1493.001	186	RTA00001073F.k.21.1	M00004167A:H04
9256	9/28/1998	1493.001	187	RTA00001072F.j.05.2	M00003865B:D10
9257	9/28/1998	1493.001	188	RTA00001063F.i.15.1	M00004335A:G05
9258	9/28/1998	1493.001	189	RTA00001062F.g.21.1	M00003907C:D02
9259	9/28/1998	1493.001	190	RTA00001073F.b.16.1	M00004027C:E06

9260	9/28/1998	1493.001	191	RTA00001062F.g.06.1	M00003895C:F05
9261	9/28/1998	1493.001	192	RTA00001071F.b.17.1	M00001360B:B01
9262	9/28/1998	1493.001	193	RTA00001073F.f.18.1	M00004087B:D05
9263	9/28/1998	1493.001	194	RTA00001074F.b.04.1	M00004280D:D10
9264	9/28/1998	1493.001	195	RTA00001072F.d.23.2	M00003762D:C02
9265	9/28/1998	1493.001	196	RTA00001073F.l.14.1	M00004173A:D03
9266	9/28/1998	1493.001	197	RTA00001061F.p.21.1	M00003747C:G12
9267	9/28/1998	1493.001	198	RTA00001071F.n.22.1	M00001598C:F02
9268	9/28/1998	1493.001	199	RTA00001073F.d.22.1	M00004059D:A09
9269	9/28/1998	1493.001	200	RTA00001072F.j.14.2	M00003876C:G11
9270	9/28/1998	1493.001	201	RTA00001071F.k.21.2	M00001528D:B12
9271	9/28/1998	1493.001	202	RTA00001074F.a.09.1	M00004269C:B10
9272	9/28/1998	1493.001	203	RTA00001073F.p.19.1	M00004253A:E02
9273	9/28/1998	1493.001	204	RTA00001061F.b.02.1	M00001358B:F12
9274	9/28/1998	1493.001	205	RTA00001063F.e.10.1	M00004240C:A06
9275	9/28/1998	1493.001	206	RTA00001074F.j.18.1	M00004427D:H04
9276	9/28/1998	1493.001	207	RTA00001073F.f.09.1	M00004084C:F05
9277	9/28/1998	1493.001	208	RTA00001071F.l.19.1	M00001558D:E02
9278	9/28/1998	1493.001	209	RTA00001073F.c.09.1	M00004036B:C11
9279	9/28/1998	1493.001	210	RTA00001074F.a.14.1	M00004270C:H05
9280	9/28/1998	1493.001	211	RTA00001074F.l.03.1	M00004466A:E04
9281	9/28/1998	1493.001	212	RTA00000611F.f.13.2	M00004175D:G10
9282	9/28/1998	1493.001	213	RTA00001074F.e.16.1	M00004343A:G07
9283	9/28/1998	1493.001	214	RTA00001073F.l.05.1	M00004170C:A12
9284	9/28/1998	1493.001	215	RTA00001074F.e.19.1	M00004347A:F10
9285	9/28/1998	1493.001	216	RTA00001073F.e.07.1	M00004067C:E05
9286	9/28/1998	1493.001	217	RTA00001062F.p.22.1	M00004142C:A06
9287	9/28/1998	1493.001	218	RTA00001061F.c.11.1	M00001382D:F03
9288	9/28/1998	1493.001	219	RTA00001062F.f.01.1	M00003877C:A08
9289	9/28/1998	1493.001	220	RTA00001072F.l.09.2	M00003893A:D03
9290	9/28/1998	1493.001	221	RTA00001072F.i.14.2	M00003847B:H01
9291	9/28/1998	1493.001	222	RTA00001063F.g.18.1	M00004295A:C02
9292	9/28/1998	1493.001	223	RTA00001062F.j.18.1	M00003977C:D01
9293	9/28/1998	1493.001	224	RTA00001061F.b.05.1	M00001360D:C12
9294	9/28/1998	1493.001	225	RTA00001074F.e.18.1	M00004344B:C06
9295	9/28/1998	1493.001	226	RTA00001061F.o.20.1	M00001677B:G01

9296	9/28/1998	1493.001	227	RTA00001062F.d.10.1	M00003822A:D02
9297	9/28/1998	1493.001	228	RTA00001062F.h.16.1	M00003919D:F01
9298	9/28/1998	1493.001	229	RTA00001063F.e.19.1	M00004251B:H12
9299	9/28/1998	1493.001	230	RTA00001061F.o.18.1	M00001675C:F05
9300	9/28/1998	1493.001	231	RTA00001072F.j.20.2	M00003879D:A09
9301	9/28/1998	1493.001	232	RTA00001071F.j.15.3	M00001485A:C04
9302	9/28/1998	1493.001	233	RTA00001071F.a.09.1	M00001340C:D09
9303	9/28/1998	1493.001	234	RTA00001074F.j.13.1	M00004423C:F03
9304	9/28/1998	1493.001	235	RTA00001071F.i.15.3	M00001466C:H11
9305	9/28/1998	1493.001	236	RTA00001071F.b.13.1	M00001358C:D09
9306	9/28/1998	1493.001	237	RTA00001061F.g.05.1	M00001441D:G02
9307	9/28/1998	1493.001	238	RTA00001063F.e.16.1	M00004249A:C09
9308	9/28/1998	1493.001	239	RTA00001072F.j.22.2	M00003880B:B08
9309	9/28/1998	1493.001	240	RTA00001063F.i.16.1	M00004335D:D03
9310	9/28/1998	1493.001	241	RTA00000611F.f.05.2	M00004174B:B12
9311	9/28/1998	1493.001	242	RTA00001071F.p.07.1	M00001631D:G08
9312	9/28/1998	1493.001	243	RTA00001071F.c.12.1	M00001375C:C11
9313	9/28/1998	1493.001	244	RTA00001074F.k.15.1	M00004450A:G07
9314	9/28/1998	1493.001	245	RTA00001061F.e.19.1	M00001419A:E01
9315	9/28/1998	1493.001	246	RTA00001073F.g.22.1	M00004108C:D07
9316	9/28/1998	1493.001	247	RTA00001061F.g.01.1	M00001437D:A12
9317	9/28/1998	1493.001	248	RTA00001072F.n.08.2	M00003923D:A03
9318	9/28/1998	1493.001	249	RTA00001074F.b.12.1	M00004286D:D02
9319	9/28/1998	1493.001	250	RTA00001061F.l.18.1	M00001576C:E03
9320	9/28/1998	1493.001	251	RTA00001074F.j.03.1	M00004414D:A01
9321	9/28/1998	1493.001	252	RTA00001072F.h.07.2	M00003824A:B11
9322	9/28/1998	1493.001	253	RTA00001072F.j.18.2	M00003877C:C11
9323	9/28/1998	1493.001	254	RTA00001063F.c.21.1	M00004198B:G08
9324	9/28/1998	1493.001	255	RTA00001073F.m.11.1	M00004181A:B05
9325	9/28/1998	1493.001	256	RTA00001061F.h.16.1	M00001463C:E12
9326	9/28/1998	1493.001	257	RTA00001073F.i.11.1	M00004128B:H11
9327	9/28/1998	1493.001	258	RTA00001062F.k.20.1	M00003997A:C08
9328	9/28/1998	1493.001	259	RTA00001062F.o.05.1	M00004101A:C12
9329	9/28/1998	1493.001	260	RTA00001073F.p.01.1	M00004237B:G01
9330	9/28/1998	1493.001	261	RTA00001072F.a.04.2	M00001647D:A02
9331	9/28/1998	1493.001	262	RTA00001073F.e.12.1	M00004071C:B06

9332	9/28/1998	1493.001	263	RTA00001073F.p.22.1	M00004253D:D04
9333	9/28/1998	1493.001	264	RTA00001072F.i.19.3	M00003853C:A09
9334	9/28/1998	1493.001	265	RTA00001071F.d.06.1	M00001386B:E01
9335	9/28/1998	1493.001	266	RTA00001073F.j.20.1	M00004149C:D11
9336	9/28/1998	1493.001	267	RTA00001074F.l.20.1	M00004502B:G05
9337	9/28/1998	1493.001	268	RTA00001072F.h.14.2	M00003829C:G07
9338	9/28/1998	1493.001	269	RTA00001062F.b.13.1	M00003788C:C05
9339	9/28/1998	1493.001	270	RTA00001061F.j.14.1	M00001514B:C02
9340	9/28/1998	1493.001	271	RTA00001072F.j.11.2	M00003870C:H03
9341	9/28/1998	1493.001	272	RTA00001074F.m.01.1	M00004507A:F11
9342	9/28/1998	1493.001	273	RTA00001063F.f.03.1	M00004264B:F03
9343	9/28/1998	1493.001	274	RTA00001071F.l.21.1	M00001559D:E02
9344	9/28/1998	1493.001	275	RTA00001072F.b.11.2	M00001669B:H04
9345	9/28/1998	1493.001	276	RTA00001074F.i.16.1	M00004406A:H12
9346	9/28/1998	1493.001	277	RTA00001061F.j.03.1	M00001500A:A02
9347	9/28/1998	1493.001	278	RTA00001062F.n.16.1	M00004085B:D12
9348	9/28/1998	1493.001	279	RTA00001073F.j.03.1	M00004140C:D04
9349	9/28/1998	1493.001	280	RTA00001072F.k.01.2	M00003880C:D06
9350	9/28/1998	1493.001	281	RTA00001074F.k.08.1	M00004445D:A04
9351	9/28/1998	1493.001	282	RTA00001062F.k.05.1	M00003985B:F06
9352	9/28/1998	1493.001	283	RTA00001073F.h.01.1	M00004109A:B07
9353	9/28/1998	1493.001	284	RTA00000611F.f.15.2	M00004176A:E07
9354	9/28/1998	1493.001	285	RTA00001073F.b.01.1	M00003995B:C06
9355	9/28/1998	1493.001	286	RTA00001072F.c.16.2	M00001694B:H12
9356	9/28/1998	1493.001	287	RTA00001073F.c.10.1	M00004036C:E10
9357	9/28/1998	1493.001	288	RTA00001062F.g.22.1	M00003908C:C04
9358	9/28/1998	1493.001	289	RTA00001074F.d.15.1	M00004323B:G12
9359	9/28/1998	1493.001	290	RTA00001061F.c.12.1	M00001383C:C04
9360	9/28/1998	1493.001	291	RTA00001073F.k.15.1	M00004165B:E03
9361	9/28/1998	1493.001	292	RTA00001072F.j.23.2	M00003880B:D03
9362	9/28/1998	1493.001	293	RTA00001073F.j.21.1	M00004150A:B09
9363	9/28/1998	1493.001	294	RTA00001073F.h.20.1	M00004123B:G05
9364	9/28/1998	1493.001	295	RTA00001063F.g.05.1	M00004285C:B06
9365	9/28/1998	1493.001	296	RTA00001061F.a.21.1	M00001352D:A09
9366	9/28/1998	1493.001	297	RTA00001061F.d.17.1	M00001399B:C04
9367	9/28/1998	1493.001	298	RTA00001072F.h.04.2	M00003819D:B02

9368	9/29/1998	1494.001	1	RTA00001082F.j.11.1	M00027137D:F05
9369	9/29/1998	1494.001	2	RTA00001082F.h.08.1	M00027042D:E02
9370	9/29/1998	1494.001	3	RTA00001082F.e.15.1	M00026936D:D01
9371	9/29/1998	1494.001	4	RTA00001082F.l.21.1	M00027204B:A08
9372	9/29/1998	1494.001	5	RTA00001082F.e.05.1	M00026910C:C05
9373	9/29/1998	1494.001	6	RTA00001082F.i.07.1	M00027085C:H12
9374	9/29/1998	1494.001	7	RTA00001082F.i.12.1	M00027096B:A01
9375	9/29/1998	1494.001	8	RTA00001082F.m.12.1	M00027218C:D06
9376	9/29/1998	1494.001	9	RTA00001082F.p.16.1	M00027364D:E08
9377	9/29/1998	1494.001	10	RTA00001082F.g.22.1	M00027028B:C12
9378	9/29/1998	1494.001	11	RTA00001069F.e.20.1	M00026857A:F02
9379	9/29/1998	1494.001	12	RTA00001082F.c.05.3	M00026811A:H01
9380	9/29/1998	1494.001	13	RTA00001083F.c.15.1	M00027529B:B11
9381	9/29/1998	1494.001	14	RTA00001082F.f.08.1	M00026964C:H02
9382	9/29/1998	1494.001	15	RTA00001082F.o.01.1	M00027280D:H01
9383	9/29/1998	1494.001	16	RTA00001082F.l.05.1	M00027190B:F06
9384	9/29/1998	1494.001	17	RTA00001082F.l.10.1	M00027196A:A10
9385	9/29/1998	1494.001	18	RTA00001069F.i.06.1	M00026972A:F04
9386	9/29/1998	1494.001	19	RTA00001082F.o.21.1	M00027339D:E10
9387	9/29/1998	1494.001	20	RTA00001069F.c.13.1	M00023390A:C04
9388	9/29/1998	1494.001	21	RTA00001069F.g.11.1	M00026914C:H10
9389	9/29/1998	1494.001	22	RTA00001082F.e.21.1	M00026945B:C10
9390	9/29/1998	1494.001	23	RTA00001083F.a.18.1	M00027396C:B06
9391	9/29/1998	1494.001	24	RTA00001069F.a.21.1	M00023298B:G07
9392	9/29/1998	1494.001	25	RTA00001083F.a.17.1	M00027393D:F01
9393	9/29/1998	1494.001	26	RTA00001083F.a.23.1	M00027439B:A09
9394	9/29/1998	1494.001	27	RTA00001083F.e.18.1	M00027642C:D11
9395	9/29/1998	1494.001	28	RTA00001083F.e.04.1	M00027618A:B08
9396	9/29/1998	1494.001	29	RTA00001069F.j.21.1	M00027067A:B02
9397	9/29/1998	1494.001	30	RTA00001082F.h.20.1	M00027069D:F02
9398	9/29/1998	1494.001	31	RTA00001069F.o.03.1	M00027386D:C02
9399	9/29/1998	1494.001	32	RTA00001082F.l.04.1	M00027189C:D04
9400	9/29/1998	1494.001	33	RTA00001082F.o.05.1	M00027282D:G01
9401	9/29/1998	1494.001	34	RTA00001069F.a.11.1	M00023284B:G06
9402	9/29/1998	1494.001	35	RTA00001069F.n.05.1	M00027283C:H12
9403	9/29/1998	1494.001	36	RTA00001069F.a.22.1	M00023299B:A01

9404	9/29/1998	1494.001	37	RTA00001069F.h.10.1	M00026942C:A06
9405	9/29/1998	1494.001	38	RTA00001082F.h.19.1	M00027067B:E09
9406	9/29/1998	1494.001	39	RTA00001082F.b.05.1	M00023343B:C08
9407	9/29/1998	1494.001	40	RTA00001082F.j.05.1	M00027131C:E07
9408	9/29/1998	1494.001	41	RTA00001083F.b.09.1	M00027459A:G12
9409	9/29/1998	1494.001	42	RTA00001082F.d.07.3	M00026871C:F12
9410	9/29/1998	1494.001	43	RTA00001083F.c.03.1	M00027499B:G02
9411	9/29/1998	1494.001	44	RTA00001082F.f.01.1	M00026949A:F04
9412	9/29/1998	1494.001	45	RTA00001082F.h.12.1	M00027053C:B06
9413	9/29/1998	1494.001	46	RTA00001082F.a.03.1	M00023282B:H09
9414	9/29/1998	1494.001	47	RTA00001082F.l.03.1	M00027188A:D12
9415	9/29/1998	1494.001	48	RTA00001082F.k.04.1	M00027154B:D05
9416	9/29/1998	1494.001	49	RTA00001069F.b.18.1	M00023340A:A10
9417	9/29/1998	1494.001	50	RTA00001069F.o.21.1	M00027546B:A11
9418	9/29/1998	1494.001	51	RTA00001082F.k.01.1	M00027152D:H06
9419	9/29/1998	1494.001	52	RTA00001083F.a.14.1	M00027388A:G05
9420	9/29/1998	1494.001	53	RTA00001069F.k.01.1	M00027085A:G10
9421	9/29/1998	1494.001	54	RTA00001069F.h.09.1	M00026941C:E11
9422	9/29/1998	1494.001	55	RTA00001069F.o.11.1	M00027462D:A12
9423	9/29/1998	1494.001	56	RTA00001083F.a.22.1	M00027438D:A03
9424	9/29/1998	1494.001	57	RTA00001082F.m.21.1	M00027231C:D08
9425	9/29/1998	1494.001	58	RTA00001083F.f.18.1	M00027752B:E05
9426	9/29/1998	1494.001	59	RTA00001082F.i.03.1	M00027083C:F06
9427	9/29/1998	1494.001	60	RTA00001082F.n.01.1	M00027234C:B05
9428	9/29/1998	1494.001	61	RTA00001082F.l.02.1	M00027184D:H02
9429	9/29/1998	1494.001	62	RTA00001082F.k.18.1	M00027178B:E04
9430	9/29/1998	1494.001	63	RTA00001069F.d.09.1	M00023413D:F04
9431	9/29/1998	1494.001	64	RTA00001069F.p.05.1	M00027607A:A09
9432	9/29/1998	1494.001	65	RTA00001069F.m.14.1	M00027231A:D01
9433	9/29/1998	1494.001	66	RTA00001083F.c.21.1	M00027557D:B06
9434	9/29/1998	1494.001	67	RTA00001069F.i.23.1	M00027023B:H12
9435	9/29/1998	1494.001	68	RTA00001082F.l.07.1	M00027193A:F07
9436	9/29/1998	1494.001	69	RTA00001082F.c.15.3	M00026850B:F07
9437	9/29/1998	1494.001	70	RTA00001082F.f.18.1	M00026982C:D08
9438	9/29/1998	1494.001	71	RTA00001082F.h.17.1	M00027062C:C04
9439	9/29/1998	1494.001	72	RTA00001082F.p.14.1	M00027363D:A08

9440	9/29/1998	1494.001	73	RTA00001069F.j.04.1	M00027028A:B06
9441	9/29/1998	1494.001	74	RTA00001069F.p.21.1	M00027740C:C05
9442	9/29/1998	1494.001	75	RTA00001082F.e.07.1	M00026913D:G11
9443	9/29/1998	1494.001	76	RTA00001082F.d.23.3	M00026905A:G11
9444	9/29/1998	1494.001	77	RTA00001083F.b.18.1	M00027484A:G03
9445	9/29/1998	1494.001	78	RTA00001069F.o.06.1	M00027396A:F07
9446	9/29/1998	1494.001	79	RTA00001082F.p.01.1	M00027343B:H05
9447	9/29/1998	1494.001	80	RTA00001082F.p.11.1	M00027356A:H02
9448	9/29/1998	1494.001	81	RTA00001083F.f.19.1	M00027759B:E11
9449	9/29/1998	1494.001	82	RTA00001082F.i.04.1	M00027083D:F06
9450	9/29/1998	1494.001	83	RTA00001082F.p.12.1	M00027357D:A02
9451	9/29/1998	1494.001	84	RTA00001082F.d.15.3	M00026882A:E07
9452	9/29/1998	1494.001	85	RTA00001082F.i.20.1	M00027115B:G04
9453	9/29/1998	1494.001	86	RTA00001069F.d.03.1	M00023401C:D12
9454	9/29/1998	1494.001	87	RTA00001082F.e.10.1	M00026928A:B06
9455	9/29/1998	1494.001	88	RTA00001082F.a.07.1	M00023295B:C03
9456	9/29/1998	1494.001	89	RTA00001069F.n.15.1	M00027329A:H04
9457	9/29/1998	1494.001	90	RTA00001082F.d.08.3	M00026872A:C10
9458	9/29/1998	1494.001	91	RTA00001083F.f.13.1	M00027728A:B03
9459	9/29/1998	1494.001	92	RTA00001082F.b.03.1	M00023340B:H12
9460	9/29/1998	1494.001	93	RTA00001069F.b.09.1	M00023321B:F06
9461	9/29/1998	1494.001	94	RTA00001082F.l.20.1	M00027202B:B09
9462	9/29/1998	1494.001	95	RTA00001083F.c.14.1	M00027528A:G03
9463	9/29/1998	1494.001	96	RTA00001069F.c.07.1	M00023369D:C05
9464	9/29/1998	1494.001	97	RTA00001083F.d.16.1	M00027598C:D06
9465	9/29/1998	1494.001	98	RTA00001069F.e.22.1	M00026858C:H05
9466	9/29/1998	1494.001	99	RTA00001082F.j.10.1	M00027137C:A03
9467	9/29/1998	1494.001	100	RTA00001069F.b.01.1	M00023301B:C01
9468	9/29/1998	1494.001	101	RTA00001069F.j.20.1	M00027066A:A04
9469	9/29/1998	1494.001	102	RTA00001069F.e.24.1	M00026861A:B05
9470	9/29/1998	1494.001	103	RTA00001069F.b.08.1	M00023321A:F07
9471	9/29/1998	1494.001	104	RTA00001069F.k.16.1	M00027131A:H02
9472	9/29/1998	1494.001	105	RTA00001069F.j.22.1	M00027072C:A11
9473	9/29/1998	1494.001	106	RTA00001069F.j.07.1	M00027036B:D07
9474	9/29/1998	1494.001	107	RTA00001083F.c.20.1	M00027551C:B07
9475	9/29/1998	1494.001	108	RTA00001069F.l.11.1	M00027169D:H06

9476	9/29/1998	1494.001	109	RTA00001069F.c.03.1	M00023363C:A04
9477	9/29/1998	1494.001	110	RTA00001069F.l.14.1	M00027175D:A05
9478	9/29/1998	1494.001	111	RTA00001083F.c.10.1	M00027518B:B07
9479	9/29/1998	1494.001	112	RTA00001082F.a.04.1	M00023287A:D08
9480	9/29/1998	1494.001	113	RTA00001069F.m.13.1	M00027225B:D03
9481	9/29/1998	1494.001	114	RTA00001082F.n.08.1	M00027250A:C04
9482	9/29/1998	1494.001	115	RTA00001069F.e.09.1	M00026819B:E02
9483	9/29/1998	1494.001	116	RTA00001082F.p.18.1	M00027369A:B03
9484	9/29/1998	1494.001	117	RTA00001082F.d.24.3	M00026906B:G03
9485	9/29/1998	1494.001	118	RTA00001069F.c.23.1	M00023398D:F10
9486	9/29/1998	1494.001	119	RTA00001069F.b.19.1	M00023340B:B07
9487	9/29/1998	1494.001	120	RTA00001082F.n.03.1	M00027237C:D04
9488	9/29/1998	1494.001	121	RTA00001069F.a.13.1	M00023289D:E06
9489	9/29/1998	1494.001	122	RTA00001069F.e.16.1	M00026846C:B01
9490	9/29/1998	1494.001	123	RTA00001069F.p.04.1	M00027603C:E02
9491	9/29/1998	1494.001	124	RTA00001069F.m.21.1	M00027248D:D01
9492	9/29/1998	1494.001	125	RTA00001082F.h.14.1	M00027056B:H07
9493	9/29/1998	1494.001	126	RTA00001069F.p.03.1	M00027592D:C05
9494	9/29/1998	1494.001	127	RTA00001069F.n.02.1	M00027266C:G12
9495	9/29/1998	1494.001	128	RTA00001082F.m.01.1	M00027209D:B09
9496	9/29/1998	1494.001	129	RTA00001083F.e.09.1	M00027628D:D08
9497	9/29/1998	1494.001	130	RTA00001069F.d.18.1	M00023432D:F09
9498	9/29/1998	1494.001	131	RTA00001069F.e.06.1	M00026810A:H04
9499	9/29/1998	1494.001	132	RTA00001069F.e.05.1	M00026809C:D10
9500	9/29/1998	1494.001	133	RTA00001083F.c.05.1	M00027502C:H02
9501	9/29/1998	1494.001	134	RTA00001069F.c.10.1	M00023373A:D01
9502	9/29/1998	1494.001	135	RTA00001082F.k.10.1	M00027164A:A09
9503	9/29/1998	1494.001	136	RTA00001083F.c.07.1	M00027507C:C06
9504	9/29/1998	1494.001	137	RTA00001082F.j.15.1	M00027142A:C01
9505	10/8/1998	1495.001	1	RTA00001079F.j.08.1	M00022217B:E03
9506	10/8/1998	1495.001	2	RTA00001081F.h.04.1	M00022854D:C04
9507	10/8/1998	1495.001	3	RTA00001078F.h.08.1	M00021624B:D03
9508	10/8/1998	1495.001	4	RTA00001079F.b.12.1	M00022056C:D12
9509	10/8/1998	1495.001	5	RTA00001066F.o.03.1	M00022074A:F05
9510	10/8/1998	1495.001	6	RTA00001067F.p.05.1	M00022640B:G10
9511	10/8/1998	1495.001	7	RTA00001079F.l.05.1	M00022260C:H07

9512	10/8/1998	1495.001	8	RTA00001078F.f.17.1	M00008083A:H11
9513	10/8/1998	1495.001	9	RTA00001079F.l.04.1	M00022259A:D04
9514	10/8/1998	1495.001	10	RTA00001079F.m.19.1	M00022368C:C11
9515	10/8/1998	1495.001	11	RTA00001081F.f.08.1	M00022831C:F11
9516	10/8/1998	1495.001	12	RTA00001079F.e.13.1	M00022113B:A12
9517	10/8/1998	1495.001	13	RTA00001081F.f.21.1	M00022838B:E05
9518	10/8/1998	1495.001	14	RTA00001079F.g.11.1	M00022152A:G05
9519	10/8/1998	1495.001	15	RTA00001067F.i.05.1	M00022392C:H06
9520	10/8/1998	1495.001	16	RTA00001067F.n.01.1	M00022561B:B09
9521	10/8/1998	1495.001	17	RTA00001080F.i.20.1	M00022569D:H03
9522	10/8/1998	1495.001	18	RTA00001081F.p.04.1	M00023096A:F03
9523	10/8/1998	1495.001	19	RTA00001078F.d.04.1	M00008023A:B03
9524	10/8/1998	1495.001	20	RTA00001080F.h.09.1	M00022546B:F12
9525	10/8/1998	1495.001	21	RTA00000631F.a.10.3	M00022362D:G11
9526	10/8/1998	1495.001	22	RTA00001078F.f.15.1	M00008082B:H10
9527	10/8/1998	1495.001	23	RTA00001078F.a.11.1	M00007948D:F08
9528	10/8/1998	1495.001	24	RTA00001078F.e.08.1	M00008052C:G11
9529	10/8/1998	1495.001	25	RTA00001078F.c.08.1	M00008012D:E07
9530	10/8/1998	1495.001	26	RTA00001078F.b.18.1	M00008001B:E11
9531	10/8/1998	1495.001	27	RTA00001078F.d.08.1	M00008023C:A06
9532	10/8/1998	1495.001	28	RTA00001080F.p.19.1	M00022711B:A05
9533	10/8/1998	1495.001	29	RTA00001078F.a.17.1	M00007965C:B02
9534	10/8/1998	1495.001	30	RTA00001078F.n.22.2	M00021958A:A04
9535	10/8/1998	1495.001	31	RTA00001079F.d.12.1	M00022090D:B03
9536	10/8/1998	1495.001	32	RTA00001078F.j.16.1	M00021696C:E02
9537	10/8/1998	1495.001	33	RTA00001080F.n.06.1	M00022655A:F09
9538	10/8/1998	1495.001	34	RTA00001067F.d.16.1	M00022214A:D01
9539	10/8/1998	1495.001	35	RTA00001078F.l.03.2	M00021865B:F06
9540	10/8/1998	1495.001	36	RTA00001080F.o.02.1	M00022684B:F11
9541	10/8/1998	1495.001	37	RTA00001067F.p.15.1	M00022652B:G06
9542	10/8/1998	1495.001	38	RTA00001079F.d.16.1	M00022094A:A09
9543	10/8/1998	1495.001	39	RTA00001068F.c.17.1	M00022826A:C08
9544	10/8/1998	1495.001	40	RTA00001080F.g.05.1	M00022527D:A09
9545	10/8/1998	1495.001	41	RTA00001081F.e.07.1	M00022813C:B09
9546	10/8/1998	1495.001	42	RTA00001066F.g.16.1	M00021653C:B06
9547	10/8/1998	1495.001	43	RTA00001066F.l.05.1	M00021972A:C10

9548	10/8/1998	1495.001	44	RTA00001066F.h.16.1	M00021691B:E04
9549	10/8/1998	1495.001	45	RTA00001081F.g.13.1	M00022844C:A01
9550	10/8/1998	1495.001	46	RTA00001067F.p.07.1	M00022641C:H03
9551	10/8/1998	1495.001	47	RTA00001080F.g.02.1	M00022525C:E09
9552	10/8/1998	1495.001	48	RTA00001080F.i.02.1	M00022559D:F10
9553	10/8/1998	1495.001	49	RTA00001080F.g.22.1	M00022541D:G06
9554	10/8/1998	1495.001	50	RTA00001067F.d.20.1	M00022216C:H02
9555	10/8/1998	1495.001	51	RTA00001079F.k.17.1	M00022252A:C01
9556	10/8/1998	1495.001	52	RTA00001068F.d.04.1	M00022838A:H05
9557	10/8/1998	1495.001	53	RTA00001079F.n.11.1	M00022377A:E02
9558	10/8/1998	1495.001	54	RTA00001066F.d.22.1	M00008053D:E09
9559	10/8/1998	1495.001	55	RTA00001068F.f.08.1	M00023002A:C02
9560	10/8/1998	1495.001	56	RTA00001081F.o.16.1	M00023038D:D04
9561	10/8/1998	1495.001	57	RTA00001080F.f.18.1	M00022518C:C04
9562	10/8/1998	1495.001	58	RTA00001080F.a.16.1	M00022434D:B06
9563	10/8/1998	1495.001	59	RTA00001080F.j.18.1	M00022590D:E08
9564	10/8/1998	1495.001	60	RTA00001080F.n.11.1	M00022659B:C01
9565	10/8/1998	1495.001	61	RTA00001078F.e.01.1	M00008048C:A08
9566	10/8/1998	1495.001	62	RTA00001078F.b.07.1	M00007992A:G04
9567	10/8/1998	1495.001	63	RTA00001078F.b.01.1	M00007985C:G07
9568	10/8/1998	1495.001	64	RTA00001080F.n.14.1	M00022664A:E04
9569	10/8/1998	1495.001	65	RTA00001078F.o.21.2	M00021980A:F03
9570	10/8/1998	1495.001	66	RTA00001078F.c.06.1	M00008012B:C05
9571	10/8/1998	1495.001	67	RTA00001080F.o.15.1	M00022695D:B02
9572	10/8/1998	1495.001	68	RTA00001080F.o.16.1	M00022696A:H03
9573	10/8/1998	1495.001	69	RTA00001081F.a.07.2	M00022720A:C01
9574	10/8/1998	1495.001	70	RTA00001078F.f.22.1	M00008089C:B08
9575	10/8/1998	1495.001	71	RTA00001078F.g.02.1	M00008093C:G08
9576	10/8/1998	1495.001	72	RTA00001078F.j.13.2	M00021689A:G05
9577	10/8/1998	1495.001	73	RTA00001078F.l.02.2	M00021864C:C07
9578	10/8/1998	1495.001	74	RTA00001078F.i.14.2	M00021667C:G10
9579	10/8/1998	1495.001	75	RTA00001079F.d.04.1	M00022087A:D01
9580	10/8/1998	1495.001	76	RTA00001079F.l.09.1	M00022263A:C01
9581	10/8/1998	1495.001	77	RTA00001067F.o.19.1	M00022627B:D01
9582	10/8/1998	1495.001	78	RTA00001068F.b.01.1	M00022714B:D04
9583	10/8/1998	1495.001	79	RTA00001079F.f.07.1	M00022128A:C05

9584	10/8/1998	1495.001	80	RTA00001068F.a.03.1	M00022669D:G07
9585	10/8/1998	1495.001	81	RTA00001066F.f.03.1	M00008088D:B01
9586	10/8/1998	1495.001	82	RTA00001067F.o.18.1	M00022627A:A02
9587	10/8/1998	1495.001	83	RTA00001079F.k.12.1	M00022249C:G09
9588	10/8/1998	1495.001	84	RTA00001081F.g.07.1	M00022843A:D02
9589	10/8/1998	1495.001	85	RTA00001079F.j.01.1	M00022214A:H05
9590	10/8/1998	1495.001	86	RTA00001067F.p.10.1	M00022648D:G11
9591	10/8/1998	1495.001	87	RTA00001081F.f.16.1	M00022836C:A07
9592	10/8/1998	1495.001	88	RTA00001080F.i.05.1	M00022561D:E06
9593	10/8/1998	1495.001	89	RTA00001067F.l.02.1	M00022490B:G12
9594	10/8/1998	1495.001	90	RTA00001068F.a.23.1	M00022709A:G02
9595	10/8/1998	1495.001	91	RTA00001067F.d.18.1	M00022214C:E09
9596	10/8/1998	1495.001	92	RTA00001066F.o.05.1	M00022077D:A12
9597	10/8/1998	1495.001	93	RTA00001066F.m.08.1	M00022015D:C11
9598	10/8/1998	1495.001	94	RTA00001066F.b.12.1	M00007978B:C04
9599	10/8/1998	1495.001	95	RTA00001066F.c.08.1	M00008002B:F09
9600	10/8/1998	1495.001	96	RTA00001081F.p.05.1	M00023096C:A03
9601	10/8/1998	1495.001	97	RTA00001081F.c.01.1	M00022746D:D05
9602	10/8/1998	1495.001	98	RTA00001079F.m.23.1	M00022370A:G07
9603	10/8/1998	1495.001	99	RTA00001079F.m.09.1	M00022300A:A05
9604	10/8/1998	1495.001	100	RTA00001081F.c.21.1	M00022785C:B10
9605	10/8/1998	1495.001	101	RTA00001079F.o.04.1	M00022383C:F05
9606	10/8/1998	1495.001	102	RTA00001080F.b.10.1	M00022449D:B05
9607	10/8/1998	1495.001	103	RTA00001078F.c.09.1	M00008012D:H04
9608	10/8/1998	1495.001	104	RTA00001078F.d.19.1	M00008044C:A05
9609	10/8/1998	1495.001	105	RTA00001081F.a.11.2	M00022722D:C07
9610	10/8/1998	1495.001	106	RTA00001080F.n.15.1	M00022664C:G10
9611	10/8/1998	1495.001	107	RTA00001078F.a.09.1	M00007941D:D07
9612	10/8/1998	1495.001	108	RTA00001078F.g.20.1	M00021614A:C09
9613	10/8/1998	1495.001	109	RTA00001066F.h.23.1	M00021841A:E11
9614	10/8/1998	1495.001	110	RTA00001081F.l.11.2	M00022922D:G06
9615	10/8/1998	1495.001	111	RTA00001079F.d.18.1	M00022096B:D10
9616	10/8/1998	1495.001	112	RTA00001066F.f.21.1	M00008100D:C08
9617	10/8/1998	1495.001	113	RTA00001078F.j.06.1	M00021680D:H08
9618	10/8/1998	1495.001	114	RTA00001067F.d.08.1	M00022205A:C02
9619	10/8/1998	1495.001	115	RTA00001068F.b.05.1	M00022717C:F05

9620	10/8/1998	1495.001	116	RTA00001079F.c.05.1	M00022071D:C08
9621	10/8/1998	1495.001	117	RTA00001078F.k.10.2	M00021852C:D12
9622	10/8/1998	1495.001	118	RTA00001081F.i.18.2	M00022884D:A07
9623	10/8/1998	1495.001	119	RTA00001066F.b.21.1	M00007996C:B11
9624	10/8/1998	1495.001	120	RTA00001066F.i.08.1	M00021851D:H06
9625	10/8/1998	1495.001	121	RTA00001068F.e.08.1	M00022915C:C09
9626	10/8/1998	1495.001	122	RTA00001079F.j.15.1	M00022220B:B06
9627	10/8/1998	1495.001	123	RTA00001078F.j.18.2	M00021698A:H03
9628	10/8/1998	1495.001	124	RTA00001066F.b.09.1	M00007977B:C11
9629	10/8/1998	1495.001	125	RTA00001079F.i.20.1	M00022207C:C01
9630	10/8/1998	1495.001	126	RTA00001080F.e.15.1	M00022506D:B03
9631	10/8/1998	1495.001	127	RTA00001080F.l.03.1	M00022617B:A01
9632	10/8/1998	1495.001	128	RTA00001080F.e.10.1	M00022501D:A09
9633	10/8/1998	1495.001	129	RTA00001067F.c.22.1	M00022184D:F07
9634	10/8/1998	1495.001	130	RTA00001081F.p.11.1	M00023097A:C03
9635	10/8/1998	1495.001	131	RTA00001081F.p.08.1	M00023096D:B11
9636	10/8/1998	1495.001	132	RTA00001080F.c.19.1	M00022471D:A05
9637	10/8/1998	1495.001	133	RTA00001081F.b.06.1	M00022736B:B03
9638	10/8/1998	1495.001	134	RTA00001081F.m.22.1	M00022983A:H04
9639	10/8/1998	1495.001	135	RTA00001081F.d.11.1	M00022801A:G04
9640	10/8/1998	1495.001	136	RTA00001081F.n.13.1	M00023002D:C12
9641	10/8/1998	1495.001	137	RTA00001067F.d.17.1	M00022214C:C11
9642	10/8/1998	1495.001	138	RTA00001081F.c.13.1	M00022772A:A06
9643	10/8/1998	1495.001	139	RTA00001078F.b.19.1	M00008001D:F11
9644	10/8/1998	1495.001	140	RTA00001078F.a.04.1	M00007931A:B07
9645	10/8/1998	1495.001	141	RTA00001078F.b.16.1	M00008000D:G11
9646	10/8/1998	1495.001	142	RTA00001078F.b.04.1	M00007987A:D10
9647	10/8/1998	1495.001	143	RTA00001078F.d.18.1	M00008044B:F07
9648	10/8/1998	1495.001	144	RTA00001068F.e.05.1	M00022904D:D04
9649	10/8/1998	1495.001	145	RTA00001078F.i.18.1	M00021674A:B07
9650	10/8/1998	1495.001	146	RTA00001066F.e.01.1	M00008054C:C03
9651	10/8/1998	1495.001	147	RTA00001078F.n.14.2	M00021949D:A05
9652	10/8/1998	1495.001	148	RTA00001067F.i.17.1	M00022413B:D07
9653	10/8/1998	1495.001	149	RTA00001079F.l.19.1	M00022278C:E04
9654	10/8/1998	1495.001	150	RTA00001081F.l.12.2	M00022923A:A09
9655	10/8/1998	1495.001	151	RTA00001067F.j.03.1	M00022420B:C08

9656	10/8/1998	1495.001	152	RTA00001068F.d.19.1	M00022898C:H07
9657	10/8/1998	1495.001	153	RTA00001081F.g.23.1	M00022853D:C05
9658	10/8/1998	1495.001	154	RTA00001081F.h.16.1	M00022860A:A07
9659	10/8/1998	1495.001	155	RTA00001079F.i.05.1	M00022192B:H07
9660	10/8/1998	1495.001	156	RTA00001068F.f.12.1	M00023012A:C06
9661	10/8/1998	1495.001	157	RTA00001067F.e.09.1	M00022235D:F07
9662	10/8/1998	1495.001	158	RTA00001066F.m.10.1	M00022018B:E09
9663	10/8/1998	1495.001	159	RTA00001080F.j.19.1	M00022591C:F03
9664	10/8/1998	1495.001	160	RTA00001080F.f.07.1	M00022513C:G04
9665	10/8/1998	1495.001	161	RTA00001080F.e.09.1	M00022500B:D01
9666	10/8/1998	1495.001	162	RTA00001080F.e.19.1	M00022509D:A12
9667	10/8/1998	1495.001	163	RTA00001066F.a.13.1	M00007948B:B07
9668	10/8/1998	1495.001	164	RTA00001079F.p.14.1	M00022407D:G07
9669	10/8/1998	1495.001	165	RTA00001079F.p.03.1	M00022399C:B02
9670	10/8/1998	1495.001	166	RTA00001079F.n.22.1	M00022381B:C12
9671	10/8/1998	1495.001	167	RTA00001078F.a.06.1	M00007937C:E08
9672	10/8/1998	1495.001	168	RTA00001078F.a.19.1	M00007973D:B03
9673	10/8/1998	1495.001	169	RTA00001078F.b.15.1	M00008000D:B06
9674	10/8/1998	1495.001	170	RTA00001079F.c.15.1	M00022078B:B04
9675	10/8/1998	1495.001	171	RTA00001079F.d.06.1	M00022088B:E05
9676	10/8/1998	1495.001	172	RTA00001067F.a.05.1	M00022118A:D08
9677	10/8/1998	1495.001	173	RTA00001078F.i.15.2	M00021668D:G09
9678	10/8/1998	1495.001	174	RTA00001066F.a.11.1	M00007947B:F07
9679	10/8/1998	1495.001	175	RTA00001078F.k.02.2	M00021846B:F05
9680	10/8/1998	1495.001	176	RTA00001066F.h.04.1	M00021669B:G02
9681	10/8/1998	1495.001	177	RTA00001066F.c.21.1	M00008015B:D08
9682	10/8/1998	1495.001	178	RTA00001080F.h.06.1	M00022544C:D08
9683	10/8/1998	1495.001	179	RTA00001067F.c.16.1	M00022177D:G02
9684	10/8/1998	1495.001	180	RTA00001080F.f.21.1	M00022522B:A05
9685	10/8/1998	1495.001	181	RTA00001080F.a.10.1	M00022425A:F11
9686	10/8/1998	1495.001	182	RTA00001081F.o.10.1	M00023034B:B10
9687	10/8/1998	1495.001	183	RTA00001078F.b.17.1	M00008001A:G11
9688	10/8/1998	1495.001	184	RTA00001078F.g.04.1	M00008094D:C02
9689	10/8/1998	1495.001	185	RTA00001080F.p.05.1	M00022704A:H08
9690	10/8/1998	1495.001	186	RTA00001067F.f.04.1	M00022256D:G11
9691	10/8/1998	1495.001	187	RTA00001066F.c.11.1	M00008003B:F09

9692	10/8/1998	1495.001	188	RTA00001081F.b.19.1	M00022743C:G05
9693	10/8/1998	1495.001	189	RTA00001081F.p.14.1	M00023097C:D10
9694	10/8/1998	1495.001	190	RTA00001067F.k.16.1	M00022467C:H07
9695	10/8/1998	1495.001	191	RTA00001081F.b.11.1	M00022737D:B02
9696	10/8/1998	1495.001	192	RTA00001080F.k.12.1	M00022601A:A09
9697	10/8/1998	1495.001	193	RTA00001066F.a.08.1	M00007943C:B02
9698	10/8/1998	1495.001	194	RTA00001081F.b.10.1	M00022737B:F12
9699	10/8/1998	1495.001	195	RTA00001080F.d.15.1	M00022488C:H02
9700	10/8/1998	1495.001	196	RTA00001079F.p.04.1	M00022399D:A07
9701	10/8/1998	1495.001	197	RTA00001067F.e.23.1	M00022251A:F07
9702	10/8/1998	1495.001	198	RTA00001068F.a.08.1	M00022684C:C12
9703	10/8/1998	1495.001	199	RTA00001078F.h.16.1	M00021628C:B09
9704	10/8/1998	1495.001	200	RTA00001081F.g.18.1	M00022848D:H09
9705	10/8/1998	1495.001	201	RTA00001081F.m.15.1	M00022968D:G06
9706	10/8/1998	1495.001	202	RTA00001067F.k.09.1	M00022459C:G05
9707	10/8/1998	1495.001	203	RTA00001080F.g.04.1	M00022527B:H05
9708	10/8/1998	1495.001	204	RTA00001081F.j.19.2	M00022902C:F11
9709	10/8/1998	1495.001	205	RTA00001081F.o.03.1	M00023023B:A05
9710	10/8/1998	1495.001	206	RTA00001079F.b.23.1	M00022067A:B03
9711	10/8/1998	1495.001	207	RTA00001078F.n.16.2	M00021951B:A01
9712	10/8/1998	1495.001	208	RTA00001067F.b.01.1	M00022134D:D12
9713	10/8/1998	1495.001	209	RTA00001080F.a.17.1	M00022435C:C05
9714	10/8/1998	1495.001	210	RTA00001080F.c.17.1	M00022469A:A05
9715	10/8/1998	1495.001	211	RTA00001068F.f.10.1	M00023003C:C10
9716	10/8/1998	1495.001	212	RTA00001081F.h.18.1	M00022861C:B04
9717	10/8/1998	1495.001	213	RTA00001066F.p.19.1	M00022106D:B06
9718	10/8/1998	1495.001	214	RTA00001080F.c.09.1	M00022464D:F12
9719	10/8/1998	1495.001	215	RTA00001078F.c.12.1	M00008014C:H01
9720	10/8/1998	1495.001	216	RTA00001080F.l.10.1	M00022622A:E08
9721	10/8/1998	1495.001	217	RTA00001078F.g.11.1	M00008099A:C12
9722	10/8/1998	1495.001	218	RTA00001068F.f.09.1	M00023003A:H01
9723	10/8/1998	1495.001	219	RTA00001067F.f.10.1	M00022261C:D06
9724	10/8/1998	1495.001	220	RTA00001080F.o.05.1	M00022687C:C11
9725	10/8/1998	1495.001	221	RTA00001078F.h.04.1	M00021620D:B06
9726	10/8/1998	1495.001	222	RTA00001078F.p.03.2	M00021981D:A11
9727	10/8/1998	1495.001	223	RTA00001080F.e.20.1	M00022510A:B09

9728	10/8/1998	1495.001	224	RTA00001078F.k.19.2	M00021861C:B08
9729	10/8/1998	1495.001	225	RTA00001078F.d.20.1	M00008045A:B05
9730	10/8/1998	1495.001	226	RTA00001078F.b.22.1	M00008006A:H02
9731	10/8/1998	1495.001	227	RTA00001068F.a.13.1	M00022701C:A05
9732	10/8/1998	1495.001	228	RTA00001080F.m.16.1	M00022641D:F08
9733	10/8/1998	1495.001	229	RTA00001080F.o.22.1	M00022702A:D10
9734	10/8/1998	1495.001	230	RTA00001080F.k.16.1	M00022604A:F06
9735	10/8/1998	1495.001	231	RTA00001067F.d.04.1	M00022199A:F09
9736	10/8/1998	1495.001	232	RTA00001067F.k.10.1	M00022460C:E12
9737	10/8/1998	1495.001	233	RTA00001078F.n.04.2	M00021931B:F04
9738	10/8/1998	1495.001	234	RTA00001078F.n.07.2	M00021945A:B04
9739	10/8/1998	1495.001	235	RTA00001081F.a.16.1	M00022725D:G05
9740	10/8/1998	1495.001	236	RTA00001078F.l.13.2	M00021879B:C11
9741	10/8/1998	1495.001	237	RTA00001078F.f.13.1	M00008082B:C05
9742	10/8/1998	1495.001	238	RTA00001079F.d.05.1	M00022087D:F12
9743	10/8/1998	1495.001	239	RTA00001067F.i.13.1	M00022406C:G03
9744	10/8/1998	1495.001	240	RTA00001068F.d.23.1	M00022902B:F10
9745	10/8/1998	1495.001	241	RTA00001078F.c.13.1	M00008014D:A11
9746	10/8/1998	1495.001	242	RTA00001078F.a.18.1	M00007969B:E10
9747	10/8/1998	1495.001	243	RTA00001068F.b.23.1	M00022765B:E03
9748	10/8/1998	1495.001	244	RTA00001078F.f.21.1	M00008085B:G01
9749	10/8/1998	1495.001	245	RTA00001067F.b.15.1	M00022144D:D09
9750	10/8/1998	1495.001	246	RTA00001078F.o.04.2	M00021963C:H04
9751	10/8/1998	1495.001	247	RTA00001081F.e.14.1	M00022817D:B09
9752	10/8/1998	1495.001	248	RTA00001078F.k.04.2	M00021847B:A09
9753	10/8/1998	1495.001	249	RTA00001079F.g.15.2	M00022158C:C08
9754	10/8/1998	1495.001	250	RTA00001067F.k.23.1	M00022477C:C07
9755	10/8/1998	1495.001	251	RTA00001079F.h.08.2	M00022176A:F02
9756	10/8/1998	1495.001	252	RTA00001078F.d.17.1	M00008028D:B01
9757	10/8/1998	1495.001	253	RTA00001067F.d.07.1	M00022203B:A05
9758	10/8/1998	1495.001	254	RTA00001068F.e.04.1	M00022903D:H02
9759	10/8/1998	1495.001	255	RTA00001068F.a.06.1	M00022682A:F10
9760	10/8/1998	1495.001	256	RTA00001078F.e.10.1	M00008054C:E07
9761	10/8/1998	1495.001	257	RTA00001079F.b.11.1	M00022056B:G12
9762	10/8/1998	1495.001	258	RTA00001066F.h.11.1	M00021676B:B12
9763	10/8/1998	1495.001	259	RTA00001079F.d.01.1	M00022084B:C03

9764	10/8/1998	1495.001	260	RTA00001067F.g.14.1	M00022363C:D03
9765	10/8/1998	1495.001	261	RTA00001066F.g.06.1	M00021625B:G07
9766	10/8/1998	1495.001	262	RTA00001081F.j.09.2	M00022893D:C06
9767	10/8/1998	1495.001	263	RTA00001068F.e.19.1	M00022963A:E07
9768	10/8/1998	1495.001	264	RTA00001079F.l.21.1	M00022282A:A11
9769	10/8/1998	1495.001	265	RTA00001078F.h.09.1	M00021624B:E11
9770	10/8/1998	1495.001	266	RTA00001078F.d.16.1	M00008027D:H09
9771	10/8/1998	1495.001	267	RTA00001079F.g.22.2	M00022167B:H02
9772	10/8/1998	1495.001	268	RTA00001066F.e.15.1	M00008075D:B01
9773	10/8/1998	1495.001	269	RTA00001080F.g.16.1	M00022538D:B02
9774	10/8/1998	1495.001	270	RTA00001080F.b.07.1	M00022447A:H06
9775	10/8/1998	1495.001	271	RTA00001078F.n.21.2	M00021958A:A03
9776	10/8/1998	1495.001	272	RTA00001078F.b.12.1	M00007998C:B04
9777	10/8/1998	1495.001	273	RTA00001066F.p.01.2	M00022099C:A10
9778	10/8/1998	1495.001	274	RTA00001066F.o.22.1	M00022095C:F03
9779	10/8/1998	1495.001	275	RTA00001080F.i.19.1	M00022568B:D03
9780	10/8/1998	1495.001	276	RTA00001079F.g.01.1	M00022138C:B07
9781	10/8/1998	1495.001	277	RTA00001079F.e.02.1	M00022102D:A10
9782	10/8/1998	1495.001	278	RTA00001079F.k.01.1	M00022233C:D11
9783	10/8/1998	1495.001	279	RTA00001079F.o.11.1	M00022386D:C04
9784	10/8/1998	1495.001	280	RTA00001068F.d.02.1	M00022834A:H02
9785	10/8/1998	1495.001	281	RTA00001078F.a.07.1	M00007939A:F06
9786	10/8/1998	1495.001	282	RTA00001081F.b.20.1	M00022743C:G06
9787	10/8/1998	1495.001	283	RTA00001067F.f.20.1	M00022273A:B03
9788	10/8/1998	1495.001	284	RTA00001079F.c.06.1	M00022072D:E12
9789	10/8/1998	1495.001	285	RTA00001068F.b.24.1	M00022768A:A10
9790	10/8/1998	1495.001	286	RTA00001080F.o.08.1	M00022691A:G01
9791	10/8/1998	1495.001	287	RTA00001078F.j.10.2	M00021687C:A04
9792	10/8/1998	1495.001	288	RTA00001080F.b.03.1	M00022444B:C04
9793	10/8/1998	1495.001	289	RTA00001067F.e.13.1	M00022240C:B03
9794	10/8/1998	1495.001	290	RTA00001081F.h.05.1	M00022856A:B09
9795	10/8/1998	1495.001	291	RTA00001067F.f.01.1	M00022252C:A04
9796	10/8/1998	1495.001	292	RTA00001080F.g.23.1	M00022542A:B06
9797	10/8/1998	1495.001	293	RTA00001080F.h.16.1	M00022548A:F02
9798	10/8/1998	1495.001	294	RTA00001080F.f.15.1	M00022517C:B01
9799	10/8/1998	1495.001	295	RTA00001080F.f.06.1	M00022513C:E10

9800	10/8/1998	1495.001	296	RTA00001081F.a.04.2	M00022716A:C01
9801	10/8/1998	1495.001	297	RTA00001078F.p.16.2	M00022001B:H10
9802	10/8/1998	1495.001	298	RTA00001081F.b.03.1	M00022734C:A03
9803	10/8/1998	1495.001	299	RTA00001080F.a.21.1	M00022441B:A06
9804	10/8/1998	1495.001	300	RTA00001079F.f.05.1	M00022127C:E01
9805	10/8/1998	1495.001	301	RTA00001080F.n.23.1	M00022681D:H10
9806	10/8/1998	1495.001	302	RTA00001078F.c.18.1	M00008016C:E06
9807	10/8/1998	1495.001	303	RTA00001068F.a.11.1	M00022697A:C08
9808	10/8/1998	1495.001	304	RTA00001068F.g.09.1	M00023095C:A09
9809	10/8/1998	1495.001	305	RTA00001068F.a.22.1	M00022709A:C01
9810	10/8/1998	1495.001	306	RTA00001079F.h.09.2	M00022176D:F05
9811	10/8/1998	1495.001	307	RTA00001079F.h.01.2	M00022169A:E11
9812	10/8/1998	1495.001	308	RTA00001078F.g.07.1	M00008097C:E04
9813	10/8/1998	1495.001	309	RTA00001078F.m.08.2	M00021908B:F03
9814	10/8/1998	1495.001	310	RTA00001080F.a.03.1	M00022417B:C01
9815	10/8/1998	1495.001	311	RTA00001079F.o.06.1	M00022384B:E06
9816	10/8/1998	1495.001	312	RTA00001079F.p.06.1	M00022401C:G07
9817	10/8/1998	1495.001	313	RTA00001078F.p.18.2	M00022001D:E06
9818	10/8/1998	1495.001	314	RTA00001068F.a.17.1	M00022705B:F08
9819	10/8/1998	1495.001	315	RTA00001078F.a.10.1	M00007948C:G01
9820	10/8/1998	1495.001	316	RTA00001079F.h.20.2	M00022184D:H07
9821	10/8/1998	1495.001	317	RTA00001081F.n.03.1	M00022986B:C02
9822	10/8/1998	1495.001	318	RTA00001080F.c.04.1	M00022460D:C07

Table 69B

SEQ ID NO:	Sample Name	Clone ID
9823	270.F5.sp6:145120	M00001401B:A02
9824	344.C4.sp6:146251	M00023363C:A04
9825	628.D9.sp6:157832	M00008028D:B01
9826	628.F7.sp6:157854	M00008023C:A06
9827	636.G12.sp6:158255	M00022077D:A12
9828	653.F3.sp6:159004	M00023284B:G06
9829	654.H6.sp6:159223	M00023369D:C05
9830	655.B2.sp6:156468	M00023413D:F04
9831	656.B11.sp6:159348	M00026905A:G11
9832	661.C10.sp6:159743	M00027169D:H06

9833	953.B04.sp6:185140	M00005434D:H02
9834	270.F5.sp6:145120	M00001401B:A02
9835	344.C4.sp6:146251	M00023363C:A04
9836	655.B2.sp6:156468	M00023413D:F04

Table 69C

SEQ ID NO:	Sequence Name	THC Accession No.
9837	RTA00001071F.i.23.3	AA173046
9838	RTA00001079F.m.19.1	THC220786
9839	RTA00001067F.i.05.1	THC233199
9840	RTA00001082F.o.01.1	THC178783
9841	RTA00001067F.n.01.1	AA173079
9842	RTA00001076F.b.13.1	AA554659
9843	RTA00001064F.p.03.1	AA432284
9844	RTA00001072F.g.05.2	H20612
9845	RTA00001064F.c.01.1	EST55879
9846	RTA00001083F.b.09.1	W30744
9847	RTA00001083F.c.03.1	THC205070
9848	RTA00001066F.h.16.1	EST14169
9849	RTA00001076F.n.10.1	THC144372
9850	RTA00001061F.e.17.1	N48670
9851	RTA00001071F.m.09.3	R56510
9852	RTA00001080F.g.02.1	THC77700
9853	RTA00001073F.i.02.2	Z46186
9854	RTA00001076F.j.14.1	THC144372
9855	RTA00001068F.d.04.1	AA011604
9856	RTA00001069F.o.11.1	AA576259
9857	RTA00001073F.k.01.1	R52934
9858	RTA00001080F.f.18.1	THC126698
9859	RTA00001075F.e.18.1	THC209874
9860	RTA00001076F.d.13.1	AA158197
9861	RTA00001065F.f.06.1	THC219476
9862	RTA00001068F.b.01.1	THC151511
9863	RTA00001068F.a.03.1	THC220020

9864	RTA00001072F.b.09.2	AA554360
9865	RTA00001076F.i.09.1	EST20991
9866	RTA00001073F.l.04.1	AA527712
9867	RTA00001067F.d.18.1	THC198501
9868	RTA00001082F.b.03.1	THC218291
9869	RTA00001082F.l.20.1	THC204015
9870	RTA00001081F.c.21.1	THC203534
9871	RTA00001069F.b.08.1	THC234347
9872	RTA00001074F.f.09.1	N53623
9873	RTA00001066F.h.23.1	THC129284
9874	RTA00001064F.h.07.1	THC161794
9875	RTA00001066F.f.21.1	T92493
9876	RTA00001069F.m.13.1	AA148143
9877	RTA00001064F.d.14.1	THC138642
9878	RTA00001068F.e.08.1	AA633643
9879	RTA00001065F.d.19.1	THC227618
9880	RTA00001069F.e.06.1	T19066
9881	RTA00001069F.e.05.1	T19066
9882	RTA00001082F.j.15.1	THC226714
9883	RTA00001067F.i.17.1	EST83778
9884	RTA00001081F.l.12.2	AA121009
9885	RTA00001080F.e.19.1	T99190
9886	RTA00001065F.d.18.2	H59526
9887	RTA00001078F.a.06.1	AA453802
9888	RTA00001065F.a.21.1	THC86626
9889	RTA00001075F.a.02.1	AA632565
9890	RTA00001066F.c.21.1	AA465322
9891	RTA00001080F.h.06.1	THC232157
9892	RTA00001067F.b.01.1	EST79811
9893	RTA00001071F.l.19.1	THC208816
9894	RTA00001062F.f.01.1	THC105335
9895	RTA00001063F.g.18.1	THC205088
9896	RTA00001062F.j.18.1	THC220715
9897	RTA00001078F.b.22.1	THC232576
9898	RTA00001064F.a.09.2	THC171312

9899	RTA00001064F.k.20.2	THC200994
9900	RTA00001080F.m.16.1	EST62430
9901	RTA00001078F.n.04.2	THC231131
9902	RTA00001071F.p.07.1	AA524115
9903	RTA00001074F.k.15.1	AA053768
9904	RTA00001073F.g.22.1	THC146930
9905	RTA00001067F.k.23.1	THC211481
9906	RTA00001068F.a.06.1	THC232664
9907	RTA00001067F.g.14.1	THC110314
9908	RTA00001072F.i.19.3	EST84170
9909	RTA00001079F.g.22.2	THC146930
9910	RTA00001061F.j.03.1	THC195525
9911	RTA00001072F.c.16.2	AA159011
9912	RTA00001061F.c.12.1	THC196151
9913	RTA00001072F.j.23.2	N99474
9914	RTA00001080F.f.06.1	R06925
9915	RTA00001080F.a.21.1	THC173393
9916	RTA00001068F.a.11.1	THC202663
9917	RTA00001078F.g.07.1	EST89489
9918	RTA00001078F.m.08.2	THC233725
9919	RTA00001068F.a.17.1	N86176

Example 46: Results of Public Database Search to Identify Function of Gene Products

SEQ ID NOS:8841-9919 were translated in all three reading frames, and the nucleotide sequences and translated amino acid sequences used as query sequences to search for homologous sequences in either the GenBank (nucleotide sequences) or Non-Redundant Protein (amino acid sequences) databases. Query and individual sequences were aligned using the BLAST 2.0 programs, available over the world wide web. (see also Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402). The sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above.

Tables 70A and 70B (inserted before the claims) provide the alignment summaries having a p value of 1 x 10^{-2} or less indicating substantial homology between the sequences of the present invention and those of the indicated public databases. Table 70A provides

the SEQ ID NO of the query sequence, the accession number of the GenBank database entry of the homologous sequence, and the p value of the alignment. Table 70A provides the SEQ ID NO of the query sequence, the accession number of the Non-Redundant Protein database entry of the homologous sequence, and the p value of the alignment. The alignments provided in Tables 70A and 70B are the best available alignment to a DNA or amino acid sequence at a time just prior to filing of the present specification. The activity of the polypeptide encoded by the SEQ ID NOS listed in Tables 70A and 70B can be extrapolated to be substantially the same or substantially similar to the activity of the reported nearest neighbor or closely related sequence. The accession number of the nearest neighbor is reported, providing a publicly available reference to the activities and functions exhibited by the nearest neighbor. The public information regarding the activities and functions of each of the nearest neighbor sequences is incorporated by reference in this application. Also incorporated by reference is all publicly available information regarding the sequence, as well as the putative and actual activities and functions of the nearest neighbor sequences listed in Table 70 and their related sequences. The search program and database used for the alignment, as well as the calculation of the p value are also indicated.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence of the corresponding polynucleotide. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of the corresponding polynucleotides.

Example 47: Identification of Contiguous Sequences Having a Polynucleotide of the Invention

The novel polynucleotides were used to screen publicly available and proprietary databases to determine if any of the polynucleotides of SEQ ID NOS:8841-9785 would facilitate identification of a contiguous sequence, *e.g.*, the polynucleotides would provide sequence that would result in 5' extension of another DNA sequence, resulting in production of a longer contiguous sequence composed of the provided polynucleotide and the other DNA sequence(s). Contiging was performed using the Gelmerge application (default settings) of GCG from the Univ. of Wisconsin.

Using these parameters, 83 contiged sequences were generated. These contiged sequences are provided as SEQ ID NOS:9800--9882 (see Table 69C). Table 69C provides the SEQ ID NO of the contig sequence, the name of the sequence used to create the contig,

and the accession number of the publicly available tentative human consensus (THC) sequence used with the sequence of the corresponding sequence name to provide the contig. The sequence name of Table 69C can be correlated with the SEQ ID NO: of the polynucleotide used to generate the contig by referring to Tables 69A and 69B.

The contiged sequences (SEQ ID NOS: 9800--9882) represent longer sequences that encompass another of the polynucleotide sequence of the invention. The contiged sequences were then translated in all three reading frames to determine the best alignment with individual sequences using the BLAST programs as described above. The sequences were masked using the XBLAST program for masking low complexity as described above. As described in more detail below, several of the contiged sequences were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (see Example 4 and Table 71 below). Thus the invention encompasses fragments, fusions, and variants of such polynucleotides that retain biological activity associated with the protein family and/or functional domain identified herein.

Example 48: Members of Protein Families

SEQ ID NOS:8841-9919 were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein family (and thus represent nmembers of these protein families) and/or comprising a known functional domain. Table 71 (inserted before claims) provides the SEQ ID NO: of the query sequence, a brief description of the profile hit, the position of the query sequence within the individual sequence (indicated as "start" and "stop"), and the orientation (Direction, "Dir") of the query sequence with respect to the individual sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Some polynucleotides exhibited multiple profile hits where the query sequence contains overlapping profile regions, and/or where the sequence contains two different functional domains. Each of the profile hits of Table 71 are described in more detail below. The acronyms for the profiles (provided in parentheses) are those used to identify the profile in the Pfam and Prosite databases. . The public information available on the Pfam and Prosite databases regarding the various profiles, including but not limited to the

activities, function, and consensus sequences of various proteins families and protein domains, is incorporated herein by reference.

Table 71

SEQ ID NO:	Profilename	Start	Stop	Direction
8937	Kazal	25	243	for
9067	helicase_C	212	389	for
9082	EFhand	275	310	for
9290	SH3	44	226	for
9313	Zincfing_C2H2	211	273	for
9345	WD_domain	80	178	for
9352	Zincfing_C2H2	147	209	for
9363	PDZ	168	395	for
9367	ras	18	395	for
9385	ANK	311	393	for
9387	Ets_Nterm	7	237	for
9446	WW_domain	120	209	for
9475	protkinase	47	400	for
9475	mkk	41	394	for
9476	trypsin	147	381	for
9480	Zincfing_C2H2	122	184	for
9533	Zincfing_CCHC	135	185	for
9561	WD_domain	18	116	for
9645	Zincfing_C3HC4	263	406	for
9758	BZIP	51	224	for
9759	Zincfing_C2H2	125	187	for
9765	FKH	9	230	for
9811	Zincfing_C2H2	202	264	for
9813	Zincfing_CCHC	262	309	for
9820	PDZ	241	468	for
9832	mkk	0	708	for
9832	protkinase	121	711	for

9835	trypsin	202	760	for
9824	trypsin	202	760	for
9858	WD_domain	18	116	for
9868	pr55	24	1293	for
9875	ATPases	74	616	for
9876	Zincfing_C2H2	122	184	for
9893	14_3_3	63	619	for
9898	helicase_C	212	448	for
9898	ATPases	59	442	for
9903	Zincfing_C2H2	211	273	for
9906	Zincfing_C2H2	125	187	for
9912	ATPases	808	1284	for
9918	protkinase	309	1022	rev
9918	neur_chan	12	508	rev
9918	Zincfing_CCHC	262	309	for
9918	Zincfing_C3HC4	557	679	for

14-3-3 Family (14_3_3; Pfam Pfam Accession No. PF00244). One SEQ ID NOcorresponds to a sequence encoding a 14-3-3 protein family member. The 14-3-3 protein family includes a group of closely related acidic homodimeric proteins of about 30 kD first identified as very abundant in mammalian brain tissues and located preferentially in neurons (Aitken et al. *Trends Biochem. Sci.* (1995) *20*:95-97; Morrison *Science* (1994) *266*:56-57; and Xiao et al. *Nature* (1995) *376*:188-191). The 14-3-3 proteins have multiple biological activities, including a key role in signal transduction pathways and the cell cycle. 14-3-3 proteins interact with kinases (*e.g.*, PKC or Raf-1), and can also function as protein-kinase dependent activators of tyrosine and tryptophan hydroxylases. The 14-3-3 protein sequences are extremely well conserved, and include two highly conserved regions: the first is a peptide of 11 residues located in the N-terminal section; the second, a 20 amino acid region located in the C-terminal section.

Ank Repeats (ANK; Pfam Accession No. PF0023). One SEQ ID NO represents a polynucleotide encoding an Ank repeat-containing protein. The ankyrin motif is a 33 amino acid sequence named after the protein ankyrin which has 24 tandem 33-amino-acid motifs. Ank repeats were originally identified in the cell-cycle-control protein cdc10

(Breeden *et al., Nature* (1987) *329*:651). Proteins containing ankyrin repeats include ankyrin, myotropin, I-kappaB proteins, cell cycle protein cdc10, the Notch receptor (Matsuno *et al., Development* (1997) *124(21)*:4265); G9a (or BAT8) of the class III region of the major histocompatibility complex (Biochem J. 290:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, and SW16. The functions of the ankyrin repeats are compatible with a role in protein-protein interactions (Bork, *Proteins* (1993) *17(4)*:363; Lambert and Bennet, *Eur. J. Biochem.* (1993) *211*:1; Kerr *et al., Current Op. Cell Biol.* (1992) *4*:496; Bennet *et al., J. Biol. Chem.* (1980) *255*:6424).

<u>ATPases Associated with Various Cellular Activities (ATPases; Pfam Accession No. PF0004).</u> Some SEQ ID NOS corresond to a sequence that encodes a member of a family of ATPases Associated with diverse cellular Activities (AAA). The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids containing an ATP-binding site (Froehlich *et al., J. Cell Biol.* (1991) *114*:443; Erdmann *et al. Cell* (1991) *64*:499; Peters *et al., EMBO J.* (1990) *9*:1757; Kunau *et al., Biochimie* (1993) *75*:209-224; Confalonieri *et al., BioEssays* (1995) *17*:639). The AAA domain, which can be present in one or two copies, acts as an ATP-dependent protein clamp (Confalonieri *et al.* (1995) *BioEssays 17*:639) and contains a highly conserved region located in the central part of the domain.

<u>Basic Region Plus Leucine Zipper Transcription Factors (BZIP; Pfam Accession No. PF00170)</u>. One SEQ ID NO represents a polynucleotide encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization.

<u>EF Hand (Efhand; Pfam Accession No. PF00036).</u> One SEQ ID NO corresponds to a polynucleotide encoding a member of the EF-hand protein family, a calcium binding domain shared by many calcium-binding proteins belonging to the same evolutionary family (Kawasaki *et al., Protein. Prof.* (1995) *2*:305-490). The domain is a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain, with a calcium ion coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand).

Ets Domain (Ets_Nterm; Pfam Accession No. PF110178). One SEQ ID NO, and thus the sequence it validates, represents a polynucleotide encoding a polypeptide with N-terminal homology in ETS domain. Proteins of this family contain a conserved domain, the "ETS-domain," that is involved in DNA binding. The domain appears to recognize purine-rich sequences; it is about 85 to 90 amino acids in length, and is rich in aromatic and positively charged residues (Wasylyk, et al., *Eur. J. Biochem.* (1993) *211*:718). The *ets* gene family encodes a novel class of DNA-binding proteins, each of which binds a specific DNA sequence and comprises an *ets* domain that specifically interacts with sequences containing the common core tri-nucleotide sequence GGA. In addition to an *ets* domain, native *ets* proteins comprise other sequences which can modulate the biological specificity of the protein. *Ets* genes and proteins are involved in a variety of essential biological processes including cell growth, differentiation and development, and three members are implicated in oncogenic process.

(FKH; Pfam Accession No.PF00250). One SEQ ID NO corresponds to a gene encoding a polypeptide comprising a forkhead domain. The forkhead domain (also known as a "winged helix") is present in a family of eukaryotic transcription factors, and is a conserved domain of about 100 amino acid residues that is involved in DNA-binding (Weigel *et al. Cell* (1990) *63*:455-456; Clark *et al. Nature* (1993) *364*:412-420). Mammalian genes that comprise a forkhead domain include those encoding: 1) transcriptional activators (*e.g.*, HNF-3-alpha, -beta, and -gamma proteins, which interact with the cis-acting regulatory regions of a number of liver genes); 2) interleukin-enhancer binding factor (ILF), which binds to purine-rich NFAT-like motifs in the HIV-1 LTR and the interleukin-2 promoter and is involved in both positive and negative regulation of important viral and cellular promoter elements; 3) transcription factor BF-1, which plays an important role in the establishment of the regional subdivision of the developing brain and in the development of the telencephalon; 4) human HTLF, which binds to the purine-rich region in human T-cell leukemia virus long terminal repeat (HTLV-I LTR); 5) transcription factors FREAC-1 (FKHL5, HFH-8), FREAC-2 (FKHL6), FREAC-3 (FKHL7, FKH-1), FREAC-4 (FKHL8), FREAC-5 (FKHL9, FKH-2, HFH-6), FREAC-6 (FKHL10, HFH-5), FREAC-7 (FKHL11), FREAC-8 (FKHL12, HFH-7), FKH-3, FKH-4, FKH-5, HFH-1 and HFH-4; 6) human AFX1 which is involved in a chromosomal translocation that causes acute leukemia; and 7) human FKHR which is involved in a chromosomal translocation that causes rhabdomyosarcoma. The fork domain is highly conserved, and is detected by two consensus patterns: the first corresponding to the N-terminal section of the domain; the second corresponding to a heptapeptide located in the central section of the domain.

Helicases conserved C-terminal domain (helicase_C; Pfam Accession No. PF00271). Some SEQ ID NOS represent polynucleotides encoding novel members of the DEAD/H helicase family. The DEAD box family comprises a number of eukaryotic and prokaryotic proteins involved in ATP-dependent, nucleic-acid unwinding. All DEAD box family members of the above proteins share a number of conserved sequence motifs, some of which are specific to the DEAD family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman, *Nature* (1988) *333*:22 and *Nature* (1988) *333*:578; http://www.expasy.ch/www/linder/HELICASES_TEXT.html). One of these motifs, called the 'D-E-A-D-box', represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be 'D-E-A-H-box' proteins (Wassarman D.A., et al., *Nature* (1991) *349*:463; Harosh I., et al., *Nucleic Acids Res.* (1991) *19*:6331; Koonin E.V., et al., *J. Gen. Virol.* (1992) *73*:989).

Kazal serine protease inhibitors family signature (Kazal; Pfam Accession No. PF00050). One SEQ ID NO corresponds to a polynucleotide of a gene encoding a serine protease inhibitor of the Kazal inhibitor family (Laskowski *et al. Annu. Rev. Biochem.* (1980) *49*:593-626). The basic structure of Kazal serine protease inhibitors such a type of inhibitor is described at Pfam Accession No. PF00050. Exemplary proteins known to belong to this family include: pancreatic secretory trypsin inhibitor (PSTI), whose physiological function is to prevent the trypsin-catalyzed premature activation of zymogens within the pancreas; mammalian seminal acrosin inhibitors; canidae and felidae submandibular gland double-headed protease inhibitors, which contain two Kazal-type domains, the first one inhibits trypsin and the second one elastase; a mouse prostatic secretory glycoprotein, induced by androgens, and which exhibits anti-trypsin activity; avian ovomucoids; chicken ovoinhibitor; and the leech trypsin inhibitor Bdellin B-3.

MAP kinase kinase (mkk). Some SEQ ID NOSrepresent members of the MAP kinase kinase (mkk) family. MAP kinases (MAPK) are involved in signal transduction, and are important in cell cycle and cell growth controls. The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast, invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates. MAPKKs are essential transducers through which signals must pass before reaching the nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) *79*:193-207; Nishida *et al.*, *Trends Biochem Sci* (1993) *18*:128-31; Ruderman *Curr Opin Cell Biol* (1993) *5*:207-13; Dhanasekaran *et al.*,

Oncogene (1998) *17*:1447-55; Kiefer *et al.*, *Biochem Soc Trans* (1997) *25*:491-8; and Hill, *Cell Signal* (1996) *8*:533-44.

Neurotransmitter-Gated Ion-Channel (neur_chan; Pfam Accession No. PF00065). One SEQ ID NO corresponds to a sequence encoding a neurotransmitter-gated ion channel. Neurotransmitter-gated ion-channels, which provide the molecular basis for rapid signal transmission at chemical synapses, are post-synaptic oligomeric transmembrane complexes that transiently form a ionic channel upon the binding of a specific neurotransmitter. Five types of neurotransmitter-gated receptors are known: 1) nicotinic acetylcholine receptor (AchR); 2) glycine receptor; 3) gamma-aminobutyric-acid (GABA) receptor; 4) serotonin 5HT3 receptor; and 5) glutamate receptor. All known sequences of subunits from neurotransmitter-gated ion-channels are structurally related, and are composed of a large extracellular glycosylated N-terminal ligand-binding domain, followed by three hydrophobic transmembrane regions that form the ionic channel, followed by an intracellular region of variable length. A fourth hydrophobic region is found at the C-terminal of the sequence.

PDZ Domain (PDZ; Pfam Accession No. PF00595.) Some SEQ ID NOScorrespond to a gene comprising a PDZ domain (also known as DHR or GLGF domain). PDZ domains comprise 80-100 residue repeats, several of which interact with the C-terminal tetrapeptide motifs X-Ser/Thr-X-Val-COO- of ion channels and/or receptors, and are found in mammalian proteins as well as in bacteria, yeast, and plants (Pontig *et al. Protein Sci* (1997) *6*(2):464-8). Proteins comprising one or more PDZ domains are found in diverse membrane-associated proteins, including members of the MAGUK family of guanylate kinase homologues, several protein phosphatases and kinases, neuronal nitric oxide synthase, and several dystrophin-associated proteins, collectively known as syntrophins (Ponting *et al. Bioessays* (1997) *19*(6):469-79). Many PDZ domain-containing proteins are localised to highly specialised submembranous sites, suggesting their participation in cellular junction formation, receptor or channel clustering, and intracellular signalling events. For example, PDZ domains of several MAGUKs interact with the C-terminal polypeptides of a subset of NMDA receptor subunits and/or with Shaker-type K+ channels. Other PDZ domains have been shown to bind similar ligands of other transmembrane receptors. In cell junction-associated proteins,the PDZ mediates the clustering of membrane ion channels by binding to their C-terminus. The X-ray crystallographic structure of some proteins comrpising PDZ domains have been solved (see, *e.g.*, Doyle *et al. Cell* (1996) *85*(7):1067-76).

Protein phosphatase 2A regulatory subunit PR55 signatures (PR55; Pfam Accession No. PF01240). One SEQ ID NO corresponds to a gene encoding a protine phosphatase 2A reguatory subunit. Protein phosphatase 2A (PP2A) is a serine/threonine phosphatase involved in many aspects of cellular function including the regulation of metabolic enzymes and proteins involved in signal transduction. PP2A is a trimeric enzyme that consists of a core composed of a catalytic subunit associated with a 65 Kd regulatory subunit (PR65), also called subunit A; this complex then associates with a third variable subunit (subunit B), which confers distinct properties to the holoenzyme (Mayer *et al. Trends Cell Biol.* (1994) *4*:287-291). One of the forms of the variable subunit is a 55 Kd protein (PR55) which is highly conserved in mammals (where three isoforms are known to exist). This subunit may perform a substrate recognition function or be responsible for targeting the enzyme complex to the appropriate subcellular compartment.

Protein Kinase (protkinase; Pfam Accession No. PF00069). Some SEQ ID NOS represent polynucleotides encoding protein kinases, which catalyze phosphorylation of proteins in a variety of pathways, and are implicated in cancer. Eukaryotic protein kinases (Hanks, *et al.*, *FASEB J.* (1995) *9*:576; Hunter, *Meth. Enzymol.* (1991) *200*:3; Hanks, *et al.*, *Meth. Enzymol.* (1991) *200*:38; Hanks, *Curr. Opin. Struct. Biol.* (1991) *1*:369; Hanks *et al.*, *Science* (1988) *241*:42) belong to a very extensive family of proteins that share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. The first region, located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, located in the central part of the catalytic domain, contains a conserved an aspartic acid residue that is important for the catalytic activity of the enzyme (Knighton, *et al.*, *Science* (1991) *253*:407).

The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the alignment in (Hanks, *et al.*, *FASEB J.* (1995) *9*:576) and covers the entire catalytic domain.

Ras family proteins (ras; Pfam Accession No. PF00071). One SEQ ID NO represents polynucleotides encoding the ras family of small GTP/GDP-binding proteins (Valencia et al., 1991, Biochemistry 30:4637-4648). Ras family members generally require a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. Among ras-related proteins, the highest degree of sequence conservation is found in four regions that are directly involved

in guanine nucleotide binding. The first two constitute most of the phosphate and Mg2+ binding site (PM site) and are located in the first half of the G-domain. The other two regions are involved in guanosine binding and are located in the C-terminal half of the molecule. Motifs and conserved structural features of the ras-related proteins are described in Valencia et al., 1991, Biochemistry 30:4637-4648.

<u>Src homology domain 3 (SH3; Pfam Accession No. PF00018).</u> One SEQ ID NO corresponds to a gene comprising a Src homology domain. The Src homology 3 (SH3) domain is a small protein domain of about 60 amino acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (e.g. Src, Abl, Lck) (Mayer *et al. Nature* (1988) *332*:272-275). Since then, it has been found in a great variety of other intracellular or membrane-associated proteins (Musacchio *et al. FEBS Lett.* (1992) *307*:55-61; Pawson *et al. Curr. Biol.* (1993) *3*:434-442; Mayer *et al. Trends Cell Biol.* (1993) *3*:8-13; Pawson *Nature* (1995) *373*:573-580). The SH3 domain has a characteristic fold which consists of five or six beta-strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices (Kuriyan *et al. Curr. Opin. Struct. Biol.* (1993) *3*:828-837). The SH3 domain is thought to mediate assembly of specific protein complexes via binding to proline-rich peptides (Morton *et al. Curr. Biol.* (1994) *4*:615-617). In general SH3 domains are found as single copies in a given protein, but there a significant number of proteins comprise two SH3 domains and a few comprise 3 or 4 copies. The profile to detect SH3 domains is based on a structural alignment consisting of 5 gap-free blocks and 4 linker regions totaling 62 match positions.

<u>Trypsin (trypsin; Pfam Accession No. PF00089)</u>. Some SEQ ID NOScorrespond to novel serine proteases of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved (Brenner *Nature* (1988) *334*:528).

<u>WD Domain, G-Beta Repeats (WD_domain; Pfam Accession No. PF00400)</u>. Some SEQ ID NOS represent a members of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) *56*:615). The alpha subunit binds to and hydrolyzes GTP; the beta and gamma subunits are required for the replacement of GDP by GTP as well as for membrane anchoring and

receptor recognition. In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta has eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat).

WW/rsp5/WWP domain signature and profile (WW domain; Pfam Accession No. PF00397). One SEQ ID NO corresponds to a gene encoding a protein comprising a WW domain. The WW domain (Bork *et al. Trends Biochem. Sci.* (1994) *19*:531-533; Andre *et al. Biochem. Biophys. Res. Commun.* (1994) *205*:1201-1205; Hofmann *et al. FEBS Lett.* (1995) *358*:153-157; Sudol *et al. FEBS Lett.* (1995) *369*:67-71 (also known as rsp5 or WWP) was discovered as a short conserved region in a number of unrelated proteins, among them dystrophin, the gene responsible for Duchenne muscular dystrophy. The domain, which spans about 35 residues, is repeated up to 4 times in some proteins. It has been shown (Chen *et al. Proc. Natl. Acad. Sci. U.S.A.* (1995) *92*:7819-7823) to bind proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and thus resembles somewhat SH3 domains. The WW domain conatins beta-strands grouped around four conserved aromatic positions, generally tryptophan. The name WW or WWP derives from the presence of two tryptophane as well as a conserved proline. The WW domain is frequently associated with other domains typical for proteins in signal transduction processes.

Zinc Finger, C2H2 Type (Zincfing_C2H2; Pfam Accession No. PF00096). Several sequences corresponded to polynucleotides encoding members of the C2H2 type zinc finger protein family, which contain zinc finger domains that facilitate nucleic acid binding (Klug *et al.*, *Trends Biochem. Sci.* (1987) *12*:464; Evans *et al.*, *Cell* (1988) *52*:1; Payre *et al.*, *FEBS Lett.* (1988) *234*:245; Miller *et al.*, *EMBO J.* (1985) *4*:1609; and Berg, *Proc. Natl. Acad. Sci. USA* (1988) *85*:99). In addition to the conserved zinc ligand residues, a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld *et al.*, *J. Biomol. Struct. Dyn.* (1993) *11*:557) The best conserved position, which is generally an aromatic or aliphatic residue, is located four residues after the second cysteine.

Zinc finger, C3HC4 type (RING finger), signature (Zincfing_C3H4; Pfam Accession No. PF00097). Some SEQ ID NOS represent polynucleotides encoding a polypeptide having a C3HC4 type zinc finger signature. A number of eukaryotic and viral proteins contain this signature, which is primarily a conserved cysteine-rich domain of 40 to 60 residues (Borden K.L.B., et al., *Curr. Opin. Struct. Biol.* (1996) *6*:395) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The

3D structure of the zinc ligation system is uniqueto the RING domain and is refered to as the "cross-brace" motif.

Zinc finger, CCHC type (Zincfing_CCHC; Pfam Accession No. PF00098). Some SEQ ID NOS correspond to genes encoding a member of the family of CCHC zinc fingers. Because the prototype CCHC type zinc finger structure is from an HIV protein, this domain is also referred to as a retrovrial-type zinc finger domain. The family also contains proteins involved in eukaryotic gene regulation, such as C. elegans GLH-1. The structure is an 18-residue zinc finger; no examples of indels in the alignment. The motif that defines a CCHC type zinc finger domain is: C-X2-C-X4-H-X4-C (Summers J Cell Biochem 1991 Jan;45(1):41-8). The domain is found in, for example, HIV-1 nucleocapsid protein, Moloney murine leukemia virus nucleocapsid protine NCp10 (De Rocquigny *et al. Nucleic Acids Res.* (1993) *21*:823-9), and myelin transcription factor 1 (Mytl) (Kim *et al. J. Neurosci. Res.* (1997) *50*:272-90).

Example 49: Differential Expression of Polynucleotides of the Invention: Description of Libraries and Detection of Differential Expression

The relative expression levels of the polynucleotides of the invention was assessed in several libraries prepared from various sources, including cell lines and patient tissue samples. Table 72 provides a summary of these libraries, including the shortened library name (used hereafter), the mRNA source used to prepared the cDNA library, the "nickname" of the library that is used in the tables below (in quotes), and the approximate number of clones in the library.

Table 72. Description of cDNA Libraries

Library (lib #)	Description	**Number of Clones in Library**
1	Human Colon Cell Line Km12 L4: High Metastatic Potential (derived from Km12C)	308731
2	Human Colon Cell Line Km12C: Low Metastatic Potential	284771
3	Human Breast Cancer Cell Line MDA-MB-231: High Metastatic Potential; micro-mets in lung	326937
4	Human Breast Cancer Cell Line MCF7: Non Metastatic	318979
8	Human Lung Cancer Cell Line MV-522: High Metastatic Potential	223620
9	Human Lung Cancer Cell Line UCP-3: Low Metastatic Potential	312503
12	Human microvascular endothelial cells (HMVEC) -	41938

Library (lib #)	Description	**Number of Clones in Library**
	UNTREATED (PCR (OligodT) cDNA library)	
13	Human microvascular endothelial cells (HMVEC) – bFGF TREATED (PCR (OligodT) cDNA library)	42100
14	Human microvascular endothelial cells (HMVEC) – VEGF TREATED (PCR (OligodT) cDNA library)	42825
15	Normal Colon - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	282722
16	Colon Tumor - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	298831
17	Liver Metastasis from Colon Tumor of UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	303467
18	Normal Colon - UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	36216
19	Colon Tumor - UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	41388
20	Liver Metastasis from Colon Tumor of UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library)	30956
21	GRRpz Cells derived from normal prostate epithelium	164801
22	WOca Cells derived from Gleason Grade 4 prostate cancer epithelium	162088
23	Normal Lung Epithelium of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library)	306198
24	Primary tumor, Large Cell Carcinoma of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library)	309349

The KM12L4, KM12C, and MDA-MB-231 cell lines are described in example 45 above. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, *e.g.*, Chandrasekaran *et al.*, *Cancer Res.* (1979) *39*:870 (MDA-MB-231 and MCF-7); Gastpar *et al.*, *J Med Chem* (1998) *41*:4965 (MDA-MB-231 and MCF-7); Ranson *et al.*, *Br J Cancer* (1998) *77*:1586 (MDA-MB-231 and MCF-7); Kuang *et al.*, *Nucleic Acids Res* (1998) *26*:1116 (MDA-MB-231 and MCF-7); Varki *et al.*, *Int J Cancer* (1987) *40*:46 (UCP-3); Varki *et al.*, *Tumour Biol.* (1990) *11*:327; (MV-522 and UCP-3); Varki *et al.*, *Anticancer Res.* (1990) *10*:637; (MV-522); Kelner *et al.*, *Anticancer Res* (1995) *15*:867 (MV-522); and Zhang *et al.*, *Anticancer Drugs* (1997) *8*:696 (MV522)). The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-

treated HMVEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMVEC were prepared by incubation with 20ng/ml VEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation. The GRRpz and WOca cell lines were provided by Dr. Donna M. Peehl, Department of Medicine, Stanford University School of Medicine. GRRpz was derived from normal prostate epithelium. The WOca cell line is a Gleason Grade 4 cell line.

Each of the libraries is composed of a collection of cDNA clones that in turn are representative of the mRNAs expressed in the indicated mRNA source. In order to facilitate the analysis of the millions of sequences in each library, the sequences were assigned to clusters. The concept of "cluster of clones" is derived from a sorting/grouping of cDNA clones based on their hybridization pattern to a panel of roughly 300 7bp oligonucleotide probes (see Drmanac *et al.*, *Genomics* (1996) *37*(1):29). Random cDNA clones from a tissue library are hybridized at moderate stringency to 300 7bp oligonucleotides. Each oligonucleotide has some measure of specific hybridization to that specific clone. The combination of 300 of these measures of hybridization for 300 probes equals the "hybridization signature" for a specific clone. Clones with similar sequence will have similar hybridization signatures. By developing a sorting/grouping algorithm to analyze these signatures, groups of clones in a library can be identified and brought together computationally. These groups of clones are termed "clusters". Depending on the stringency of the selection in the algorithm (similar to the stringency of hybridization in a classic library cDNA screening protocol), the "purity" of each cluster can be controlled. For example, artifacts of clustering may occur in computational clustering just as artifacts can occur in "wet-lab" screening of a cDNA library with 400 bp cDNA fragments, at even the highest stringency. The stringency used in the implementation of cluster herein provides groups of clones that are in general from the same cDNA or closely related cDNAs. Closely related clones can be a result of different length clones of the same cDNA, closely related clones from highly related gene families, or splice variants of the same cDNA.

Differential expression for a selected cluster was assessed by first determining the number of cDNA clones corresponding to the selected cluster in the first library (Clones in 1st), and the determining the number of cDNA clones corresponding to the selected cluster in the second library (Clones in 2nd). Differential expression of the selected cluster in the first library relative to the second library is expressed as a "ratio" of percent expression between the two libraries. In general, the "ratio" is calculated by: 1) calculating the percent

expression of the selected cluster in the first library by dividing the number of clones corresponding to a selected cluster in the first library by the total number of clones analyzed from the first library; 2) calculating the percent expression of the selected cluster in the second library by dividing the number of clones corresponding to a selected cluster in a second library by the total number of clones analyzed from the second library; 3) dividing the calculated percent expression from the first library by the calculated percent expression from the second library. If the "number of clones" corresponding to a selected cluster in a library is zero, the value is set at 1 to aid in calculation. The formula used in calculating the ratio takes into account the "depth" of each of the libraries being compared, *i.e.*, the total number of clones analyzed in each library.

In general, a polynucleotide is said to be significantly differentially expressed between two samples when the ratio value is greater than at least about 2, preferably greater than at least about 3, more preferably greater than at least about 5 , where the ratio value is calculated using the method described above. The significance of differential expression is determined using a z score test (Zar, Biostatistical Analysis, Prentice Hall, Inc., USA, “Differences between Proportions,” pp 296-298 (1974).

Examples 50-54: Differential Expression of Polynucleotides of the Invention

A number of polynucleotide sequences have been identified that are differentially expressed between, for example, cells derived from high metastatic potential cancer tissue and low metastatic cancer cells, and between cells derived from metastatic cancer tissue and normal tissue. Evaluation of the levels of expression of the genes corresponding to these sequences can be valuable in diagnosis, prognosis, and/or treatment (*e.g.*, to facilitate rationale design of therapy, monitoring during and after therapy, *etc.*). Moreover, the genes corresponding to differentially expressed sequences described herein can be therapeutic targets due to their involvement in regulation (*e.g.*, inhibition or promotion) of development of, for example, the metastatic phenotype. For example, sequences that correspond to genes that are increased in expression in high metastatic potential cells relative to normal or non-metastatic tumor cells may encode genes or regulatory sequences involved in processes such as angiogenesis, differentiation, cell replication, and metastasis.

Detection of the relative expression levels of differentially expressed polynucleotides described herein can provide valuable information to guide the clinician in the choice of therapy. For example, a patient sample exhibiting an expression level of one or more of these polynucleotides that corresponds to a gene that is increased in expression

in metastatic or high metastatic potential cells may warrant more aggressive treatment for the patient. In contrast, detection of expression levels of a polynucleotide sequence that corresponds to expression levels associated with that of low metastatic potential cells may warrant a more positive prognosis than the gross pathology would suggest.

A number of polynucleotide sequences of the present invention are differentially expressed between human microvascular endothelial cells (HMVEC) that have been treated with growth factors relative to untreated HMVEC. Sequences that are differentially expressed between growth factor-treated HMVEC and untreated HMVEC can represent sequences encoding gene products involved in angiogenesis, metastasis (cell migration), and other development and oncogenic processes. For example, sequences that are more highly expressed in HMVEC treated with growth factors (such as bFGF or VEGF) relative to untreated HMVEC can serve as drug targets for chemotherapeutics, *e.g.*, decreasing expression of such up-regulated genes or inhibitin gthe activity of the encoded gene product would serve to inhibit tumor cell angiogenesis. Detection of expression of these sequences in colon cancer tissue can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. A patient sample displaying an increased level of one or more of these polynucleotides may thus warrant closer attention or more frequent screening procedures to catch the malignant state as early as possible.

The differential expression of the polynucleotides described herein can thus be used as, for example, diagnostic markers, prognostic markers, for risk assessment, patient treatment and the like. These polynucleotide sequences can also be used in combination with other known molecular and/or biochemical markers. The following examples provide relative expression levels of polynucleotides from specified cell lines and patient tissue samples.

Example 50: High Metastatic Potential Breast Cancer Versus Low Metastatic Breast Cancer Cells

The tables bellow summarize the data for polynucleotides that represent genes differentially expressed between high metastatic potential and low metastatic potential breast cancer cells.

Table 73. High metastatic potential breast (lib3) > low metastatic potential breast cancer cells (lib4)

SEQ ID NO:	Lib 3 Clones	Lib4 Clones	Lib3/Lib4
9621	13	0	12.68
9618	9	0	8.78
9596	8	0	7.81
9619	7	0	6.83
9531	7	0	6.83
9526	7	0	6.83
9756	6	0	5.85

Table 74. Low metastatic potential breast (lib4) > high metastatic potential breast cancer cells (lib3)

Table 74			
SEQ ID NO:	Lib 3 Clones	Lib4 Clones	Lib4/Lib3
9398	0	340	348.48
9496	0	64	65.6
9501	0	57	58.42
9487	0	43	44.07
9387	0	41	42.02
9488	0	40	41
9432	4	115	29.47
9494	0	28	28.7
9486	0	21	21.52
9476	3	61	20.84
9373	1	17	17.42
9389	0	17	17.42
9490	3	50	17.08
9429	0	16	16.4
8950	0	16	16.4
9497	0	16	16.4
9464	0	16	16.4
9477	0	13	13.32
9376	0	12	12.3
9493	1	11	11.27
9402	1	11	11.27
9427	1	11	11.27
9449	1	11	11.27
9430	0	10	10.25
9481	0	10	10.25
9372	1	10	10.25
9463	0	9	9.22
9431	0	8	8.2
9361	0	8	8.2
9054	0	7	7.17
9447	0	7	7.17
9394	0	7	7.17
9395	0	7	7.17
9422	0	7	7.17
9424	0	7	7.17
9439	0	7	7.17

Table 74			
SEQ ID NO:	**Lib 3 Clones**	**Lib4 Clones**	**Lib4/Lib3**
9401	0	6	6.15
9412	0	6	6.15
9199	0	6	6.15
9475	0	6	6.15
8953	0	6	6.15
9443	0	6	6.15

Example 51: High Metastatic Potential Lung Cancer Versus Low Metastatic Lung Cancer Cells

The following summarizes polynucleotides that represent genes differentially expressed between high metastatic potential lung cancer cells and low metastatic potential lung cancer cells:

Table 75. High metastatic potential lung (lib8) > low metastatic potential lung cancer cells (lib9)

SEQ ID NO:	Lib 8 Clones	Lib 9 Clones	Lib8/Lib9
9411	35	1	48.91
9809	8	0	11.18
9190	5	0	6.99

Example 52: High Metastatic Potential Colon Cancer Versus Low Metastatic Colon Cancer Cells

Table 76 summarizes polynucleotides that represent genes differentially expressed between high metastatic potential and low metastatic potential colon cancer cells:

Table 76. Low metastatic potential colon (lib2) > high metastatic potential colon cancer cells (lib1)

SEQ ID NO:	Lib1 Clones	Lib2 Clones	Lib2/Lib1
8897	0	8	8.67
8943	0	6	6.5
9029	0	6	6.5

Example 53: High Tumor Potential Colon Tissue Vs. Metastasized Colon Cancer Tissue

The following table summarizes polynucleotides that represent genes differentially expressed between high tumor potential colon cancer cels and cells derived from high metastatic potential colon cancer cells of a patient.

Table 77. High tumor potential colon tissue (lib16) vs. high metastatic colon tissue (lib17)

SEQ ID NO:	Lib 16	Lib 17	Lib17/Lib16
8940	0	7	6.89
9210	3	12	3.94

Example 54: Differential Expression Across Multiple Libraries

A number of polynucleotide sequences have been identified that represent genes that are differentially expressed across multiple libraries. Expression of these sequences in a tissue or any origin can be valuable in determining diagnostic, prognostic and/or treatment information associated with the prevention of achieving the malignant state in these tissues, and can be important in risk assessment for a patient. These polynucleotides can also serve as non-tissue specific markers of, for example, risk of metastasis of a tumor. The differential expression data for these sequences is provided in Table 78 below.

Table 78. Genes Differentially Expressed Across Multiple Library Comparisons

SEQ ID NO:	Cell or Tissue Sample and Cancer State Compared	RATIO
8874	Low Met Colon (lib2) > High Met Colon (lib1)	8.67
8874	High Met Breast (lib3) > Low Met Breast (Lib4)	5.85
9049	Low Met Lung (lib9) > High Met Lung (lib8)	17.44
9049	Colon Tumor Tissue (lib16) > Normal Colon Tissue (lib15)	3.42
9049	Colon Tumor Tissue (lib19) > Normal Colon Tissue (lib18)	66.5
9049	High Met Colon Tissue (lib20) > Normal Colon Tissue (lib18)	14.04
9049	Colon Tumor Tissue (lib19) > High Met Colon Tissue (lib20)	4.74
9156	High Met Colon (lib1) > Low Met Colon (lib2)	5.76
9156	Low Met Breast (lib4) > High Met Breast (Lib3)	17.28
9485	Low Met Breast (lib4) > High Met Breast (Lib3)	6.15
9485	High Met Lung (lib8) > Low Met Lung (lib9)	19.56
9694	High Met Breast (lib3) > Low Met Breast (Lib4)	9.76
9694	HMVEC-bFGF (lib13) > HMVEC (lib12)	4.98
9694	Lung Tumor Tissue (lib24) > Normal Lung Tissue (lib23)	5.94

Key for Table 78: High Met = high metastatic potential; Low Met = low metastatic potential; met = metastasized; tumor = non-metastasized tumor; HMVEC = human microvascular endothelial cell; bFGF = bFGF treated.

Detection of expression of genes that correspond to the above polynucleotides may be of particular interest in diagnosis, prognosis, risk assesment, and monitoring of treatment. Furthermore, differential expression of a specific gene across multiple libraries can also be indicative of a gene whose expression is associated with, for example, suppression of the metastatic phenotype or with development of the cell toward a metastatic phenotype. For example, SEQ ID NO:9012 corresponds to a gene that is expressed at relatively higher levels in colon tumor tissue than in high metastatic potential colon tumor tissue, and at relatively higher levels in high metastatic potential colon tumor tissue than in normal colon tissue. Thus a relatively increased level of expression of the gene corresponding to SEQ ID NO:9012 may be used as marker of a pre-metastatic colon cells either alone or in combination with other markers.

Some polynucleotides exhibited opposite differential expression trends in libraries of different origin (see, *e.g.*., SEQ ID NO:9119). These data suggest that the differential expressio patterns of some gene associated with development of metastases indicate a unique role for those genes specific for the tissue of origin.

Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

Deposit Information. The following materials were deposited with the American Type Culture Collection (CMCC = Chiron Master Culture Collection).

Table 79. Cell Lines Deposited with ATCC

Cell Line	**Deposit Date**	**ATCC Accession No.**	**CMCC Accession No.**

KM12L4-A	March 19, 1998	CRL-12496	11606
Km12C	May 15, 1998	CRL-12533	11611
MDA-MB-231	May 15, 1998	CRL-12532	10583
MCF-7	October 9, 1998	CRL-12584	10377

In addition, pools of selected clones, as well as libraries containing specific clones, were assigned an "ES" number (internal reference) and deposited with the ATCC. Table 80 below provides the ATCC Accession Nos. of the ES deposits, all of which were deposited on or before May 13, 1999. The names of the clones contained within each of these deposits are provided in the tables 81 and 82.

Table 80: Pools of Clones and Libraries Deposited with ATCC on or before September 23, 1999

Library No.	CMCC No.	ATCC Deposit No.	Library No.	CMCC No.	ATCC Deposit No.
ES55	5058	PTA-739	ES65	5068	PTA-749
ES56	5059	PTA-740	ES66	5069	PTA-750
ES57	5060	PTA-741	ES67	5070	PTA-751
ES58	5061	PTA-742	ES68	5071	PTA-752
ES59	5062	PTA-743	ES69	5072	PTA-753
ES60	5063	PTA-744	ES70	5073	PTA-754
ES61	5064	PTA-745	ES71	5074	PTA-755
ES62	5065	PTA-746	ES72	5075	PTA-756
ES63	5066	PTA-747	ES73	5076	PTA-757
ES64	5067	PTA-748	ES74	5077	PTA-758

Table 81			
ES55	**ES56**	**ES57**	**ES58**
M00004170C:H06	M00004036B:C11	M00004288D:E07	M00023298B:G07
M00004170D:C06	M00004064B:G03	M00004318D:D07	M00026819B:E02
M00004171D:H10	M00004067C:E05	M00004356C:D02	M00026914C:H10
M00004174B:B12	M00004099C:F04	M00004391C:F12	M00027023B:H12
M00004175D:G10	M00004103A:E06	M00004386C:C03	M00027085A:G10
M00004176A:E07	M00004128B:H11	M00004414D:C11	M00027248D:D01
M00001352D:A09	M00004167A:H04	M00004422C:A01	M00027546B:A11
M00001345C:B10	M00004158C:B01	M00004427D:H04	M00023299B:A01
M00001382D:F03	M00004165B:E03	M00004502B:G05	M00026857A:F02
M00001419A:E01	M00004181A:B05	M00004495D:A05	M00026858C:H05
M00001437D:A12	M00003993C:G11	M00005364C:A02	M00026861A:B05
M00001441D:G02	M00004046C:A04	M00005375B:H03	M00026846C:B01
M00001601D:A03	M00004034A:G03	M00005420C:E10	M00027131A:H02
M00001677B:G01	M00004036C:E10	M00005413B:B02	M00027396A:F07

M00001678A:B10	M00004043C:A06	M00005438D:A08	M00023301B:C01
M00001675C:F05	M00004067C:C10	M00005453B:B06	M00023321B:F06
M00001360D:C12	M00004068A:A03	M00005446B:D10	M00023401C:D12
M00001389C:E01	M00004069A:E04	M00005493D:H12	M00026941C:E11
M00001390C:H05	M00004071C:B06	M00005476D:A11	M00027067A:B02
M00001399B:C04	M00004127C:C08	M00005482A:D08	M00027036B:D07
M00001507A:H06	M00004157C:E06	M00005485C:F09	M00027329A:H04
M00003747C:G12	M00004165D:H12	M00005563C:D05	M00027740C:C05
M00001358B:F12	M00003995B:C06	M00005569B:E04	M00023340A:A10
M00001360B:F09	M00004090A:B11	M00005621B:C09	M00026942C:A06
M00001392A:F02	M00004084C:F05	M00005628D:A10	M00027066A:A04
M00001397D:G04	M00004087A:H06	M00005629B:G06	M00027072C:A11
M00001463C:E12	M00004110A:G03	M00004866C:H08	M00027028A:B06
M00001531B:A03	M00004117D:F06	M00004872C:G03	M00023282B:H09
M00001507D:F09	M00004150A:B09	M00005358B:D10	M00023295B:C03
M00001513B:F05	M00004140C:D04	M00005385D:B08	M00026811A:H01
M00001514B:C02	M00004175D:D05	M00005392C:B03	M00026850B:F07
M00001576C:E03	M00004176A:H05	M00005395C:C11	M00026913D:G11
M00003756D:B09	M00004170C:A12	M00005396A:C01	M00026936D:D01
M00003907C:D02	M00004237B:G01	M00005435B:F01	M00027083C:F06
M00003926A:D01	M00004253A:E02	M00005464B:B08	M00027152D:H06
M00003928D:A04	M00003997D:G03	M00005505B:D10	M00027209D:B09
M00003935D:E04	M00003998C:D04	M00005509D:G05	M00027339D:E10
M00003985B:F06	M00004027C:E06	M00005614A:B07	M00027282D:G01
M00004063B:B12	M00004059D:A09	M00005721C:A12	M00023287A:D08
M00004101A:C12	M00004087B:D05	M00005705D:G09	M00026928A:B06
M00004104C:F06	M00004114C:B09	M00005709D:H05	M00027028B:C12
M00004107A:E02	M00004140B:C02	M00004859D:D01	M00027115B:G04
M00004108B:D04	M00004149C:D11	M00005342D:E04	M00027096B:A01
M00003856A:H10	M00004168D:F05	M00005363D:C05	M00027154B:D05
M00003908C:C04	M00004176B:H09	M00005353C:H01	M00027164A:A09
M00003895C:F05	M00004173A:D03	M00005386C:G01	M00027218C:D06
M00003939B:C02	M00004209B:G01	M00005388B:B02	M00023343B:C08
M00003997A:C08	M00004253D:D04	M00005396C:H04	M00026871C:F12
M00004066D:C02	M00004275A:H07	M00005434A:F11	M00026882A:E07
M00004105C:C05	M00004269C:B10	M00005434C:E02	M00027067B:E09

M00003788B:C08	M00004298A:H09	M00005473C:F02	M00027062C:C04
M00003788C:C05	M00004347A:F10	M00005459B:A01	M00027131C:E07
M00003835B:C05	M00004337A:A07	M00005469A:D10	M00027137D:F05
M00003820B:G04	M00004372A:A08	M00005505D:H08	M00027204B:A08
M00003888C:G08	M00004406D:E11	M00005509B:E10	M00027188A:D12
M00003977D:H04	M00004449B:B05	M00005616B:E11	M00027190B:F06
M00004029D:H03	M00004507A:F11	M00005589B:H12	M00027193A:F07
M00004034A:A05	M00004276A:C06	M00005721D:B03	M00022362D:G11
M00004140D:E03	M00004270C:H05	M00005698A:H12	M00007947B:F07
M00003775C:C01	M00004343A:G07	M00006613C:C02	M00007948B:B07
M00003776B:F08	M00004344B:C06	M00006617A:A06	M00008003B:F09
M00003839D:C03	M00004373D:G10	M00006584D:D01	M00008054C:C03
M00003818C:D02	M00004368A:G11	M00006594B:D05	M00008075D:B01
M00003820C:E08	M00004371B:A05	M00006600D:G07	M00022074A:F05
M00003822A:D02	M00004403A:A02	M00006631D:G09	M00007943C:B02
M00003877C:G01	M00004445D:A04	M00006635A:C01	M00008002B:F09
M00003880A:G10	M00004447A:A10	M00006726D:H10	M00021653C:B06
M00003919D:F01	M00004603D:D09	M00006874D:E01	M00021851D:H06
M00003960D:E09	M00004326D:D06	M00006882C:D03	M00022015D:C11
M00004081A:E11	M00004323B:G12	M00006925B:B02	M00022018B:E09
M00004085B:D12	M00004350A:C04	M00006946B:C08	M00022095C:F03
M00004142C:A06	M00004357A:B10	M00006949B:C07	M00007996C:B11
M00004135D:D01	M00004360B:B08	M00007026A:A03	M00007977B:C11
M00004198B:G08	M00004385D:D06	M00006712A:F01	M00008088D:B01
M00004185B:H03	M00004414D:A01	M00006727A:H12	M00021676B:B12
M00004187A:B05	M00004415A:A01	M00006815D:D11	M00021972A:C10
M00004251B:H12	M00004423A:B05	M00006805D:H12	M00022099C:A10
M00004232D:G11	M00004423C:F03	M00006934B:B11	M00022106D:B06
M00004240A:D03	M00004426B:H06	M00007019B:G01	M00007978B:C04
M00004285C:B06	M00004504C:G07	M00007038D:D01	M00008053D:E09
M00004292A:C08	M00004466A:E04	M00007041C:C05	M00021669B:G02
M00004335A:G05	M00004498D:A11	M00006630A:E05	M00022118A:D08
M00004240C:A06	M00004292A:F03	M00006623C:G07	M00022251A:F07
M00004249A:C09	M00004280D:D10	M00006694D:G06	M00022235D:F07
M00004335D:D03	M00004286D:D02	M00006668D:B10	M00022240C:B03
M00004378A:H10	M00004870D:E05	M00006688A:F09	M00022406C:G03

M00004381A:E10	M00004871C:C04	M00006745B:C05	M00022459C:G05
M00004444C:H11	M00004872A:D07	M00006846A:B03	M00022627B:D01
M00004225A:E03	M00005395D:D11	M00006823A:H06	M00022184D:F07
M00004284A:C09	M00005395D:B12	M00006925A:B09	M00022177D:G02
M00004264B:F03	M00005412D:G07	M00006894D:A07	M00022460C:E12
M00004404C:B03	M00005413D:G12	M00006895D:A02	M00022627A:A02
M00004410A:F06	M00005513A:H01	M00006991B:E05	M00022144D:D09
M00004412A:G05	M00005515D:G02	M00006994A:C12	M00022203B:A05
M00001340C:A08	M00005607A:C08	M00007046D:E10	M00022214C:C11
M00001340C:D09	M00005366D:E12	M00006577A:B01	M00022252C:A04
M00001395D:B04	M00005618C:H11	M00006630A:E09	M00022420B:C08
M00001466C:H11	M00005708C:D11	M00006619A:G11	M00022640B:G10
M00001528D:B12	M00005810B:C07	M00006704A:C11	M00022641C:H03
M00001517C:A10	M00006795C:B12	M00022127C:E01	M00022652B:G06
M00001561A:G10	M00006755C:C03	M00022128A:C05	M00022216C:H02
M00001565C:F06	M00006756D:G07	M00022176D:F05	M00022199A:F09
M00001569A:H01	M00006779D:F03	M00022214A:H05	M00022214A:D01
M00001341A:H10	M00004821D:C03	M00022220B:B06	M00022273A:B03
M00001375C:C11	M00005358A:H03	M00022278C:E04	M00022256D:G11
M00001397C:F01	M00005480C:A04	M00022282A:A11	M00022261C:D06
M00001431A:F03	M00005481C:H05	M00022260C:H07	M00022490B:G12
M00001457D:E08	M00005490B:B02	M00022263A:C01	M00022648D:G11
M00001505C:C10	M00005820A:H11	M00022377A:E02	M00022709A:G02
M00001615A:D01	M00006621B:B06	M00022399C:B02	M00022701C:A05
M00001618C:E01	M00006752C:D04	M00022056C:D12	M00022826A:C08
M00001358C:D09	M00006757D:H04	M00022087A:D01	M00022963A:E07
M00001360B:B01	M00005000A:H05	M00022088B:E05	M00022904D:D04
M00001391C:B05	M00005296D:G03	M00022090D:B03	M00023095C:A09
M00001389B:B12	M00005378B:B04	M00022094A:A09	M00022684C:C12
M00001485A:C04	M00005461C:D11	M00022096B:D10	M00022765B:E03
M00001559D:E02	M00005464D:D07	M00022176A:F02	M00022898C:H07
M00001545D:F12	M00005657B:F11	M00022217B:E03	M00022902B:F10
M00001549C:F10	M00006596D:H02	M00022259A:D04	M00023003A:H01
M00001579C:E07	M00005826B:F10	M00022381B:C12	M00022768A:A10
M00001630A:E08	M00006577B:F01	M00022399D:A07	M00022834A:H02
M00001386B:E01	M00006582A:F12	M00022401C:G07	M00023002A:C02

M00001389A:F03	M00006664A:C05	M00022407D:G07	M00023003C:C10
M00001418C:F06	M00006678C:B07	M00022417B:C01	M00023012A:C06
M00001454D:H09	M00006840A:A12	M00022435C:C05	M00007973D:B03
M00001442D:D09	M00005020B:D10	M00022471D:A05	M00007939A:F06
M00001450D:H12	M00005296B:H07	M00022464D:F12	M00007941D:D07
M00001479D:B10	M00005403A:D12	M00022469A:A05	M00007948D:F08
M00001598C:F02	M00005376B:E08	M00022500B:D01	M00008012D:H04
M00001594A:H01	M00005378C:B12	M00022506D:B03	M00008014D:A11
M00001657D:D07	M00005397A:G08	M00022542A:B06	M00008048C:A08
M00003772C:F12	M00005449D:D04	M00022527D:A09	M00008099A:C12
M00003844D:B02	M00005465A:A07	M00022568B:D03	M00021668D:G09
M00003845B:A04	M00005648C:C11	M00022561D:E06	M00021861C:B08
M00003845C:F08	M00006595C:B08	M00022687C:C11	M00021980A:F03
M00003848A:E08	M00006816D:D08	M00022695D:B02	M00007931A:B07
M00003880C:D06	M00006835D:C08	M00022425A:F11	M00007948C:G01
M00001647D:A02	M00006914C:D07	M00022434D:B06	M00007969B:E10
M00001655C:F07	M00007177A:G07	M00022460D:C07	M00008012B:C05
M00003804D:F12	M00006920B:H07	M00022510A:B09	M00008012D:E07
M00003884C:G09	M00007161C:D12	M00022501D:A09	M00008014C:H01
M00003916D:A10	M00006968D:H02	M00022541D:G06	M00008016C:E06
M00003943B:C12	M00006936C:G11	M00022527B:H05	M00008052C:G11
M00003935A:C04	M00006945D:A07	M00022538D:B02	M00008054C:E07
M00003937D:F09	M00007047C:H04	M00022559D:F10	M00008093C:G08
M00001683B:F12	M00007065D:A03	M00022569D:H03	M00021614A:C09
M00001669B:H04	M00007079D:H01	M00022601A:A09	M00008094D:C02
M00003762D:C02	M00006968A:H05	M00022604A:F06	M00021667C:G10
M00003788D:E06	M00007078B:H04	M00022684B:F11	M00021674A:B07
M00003824A:B11	M00007186A:A12	M00022702A:D10	M00021846B:F05
M00003865B:D10	M00004852B:H08	M00022691A:G01	M00021847B:A09
M00003870C:H03	M00005382A:G09	M00022696A:H03	M00021963C:H04
M00003901B:C02	M00005418C:B09	M00022444B:C04	M00007985C:G07
M00003893A:D03	M00005420C:E03	M00022447A:H06	M00008001D:F11
M00003931A:G01	M00005450C:G09	M00022488C:H02	M00007992A:G04
M00003973A:D09	M00005444D:D01	M00022522B:A05	M00008000D:B06
M00001660A:B10	M00005494C:F08	M00022513C:G04	M00008001A:G11
M00003761C:C05	M00005479C:A05	M00022517C:B01	M00008044C:A05

M00003829C:G07	M00005486A:F07	M00022546B:F12	M00008085B:G01
M00003833D:F11	M00005538C:H11	M00022591C:F03	M00008082B:C05
M00003879D:A09	M00005648C:E10	M00022617B:A01	M00008083A:H11
M00003880B:B08	M00005621A:B05	M00022681D:H10	M00021624B:E11
M00003861D:G10	M00004847D:G01	M00022659B:C01	M00021689A:G05
M00003876C:G11	M00005342B:G01	M00022664C:G10	M00021865B:F06
M00003877C:C11	M00005305A:H01	M00022711B:A05	M00021879B:C11
M00003902C:D02	M00026906B:G03	M00022704A:H08	M00021958A:A03
M00003933A:B04	M00026872A:C10	M00022449D:B05	M00021945A:B04
M00003923D:A03	M00026964C:H02	M00022548A:F02	M00021981D:A11
M00003989D:A02	M00026982C:D08	M00022590D:E08	M00007987A:D10
M00003991A:D05	M00027069D:F02	M00022622A:E08	M00007998C:B04
M00004030C:E05	M00027042D:E02	M00022655A:F09	M00008001B:E11
M00004048A:E10	M00027056B:H07	M00022664A:E04	M00008045A:B05
M00006680D:A01	M00027137C:A03	M00022720A:C01	M00008023A:B03
M00006688C:C12	M00027184D:H02	M00022722D:C07	M00008027D:H09
M00006740A:A06	M00027189C:D04	M00022746D:D05	M00008044B:F07
M00006757A:C09	M00027196A:A10	M00022772A:A06	M00008089C:B08
M00006859D:E11	M00027357D:A02	M00022813C:B09	M00021620D:B06
M00006917B:C05	M00027369A:B03	M00022853D:C05	M00021624B:D03
M00006919A:H12	M00027439B:A09	M00022843A:D02	M00021628C:B09
M00006993B:F02	M00027393D:F01	M00022844C:A01	M00021680D:H08
M00007093C:C11	M00027557D:B06	M00022968D:G06	M00021687C:A04
M00007047D:C02	M00027502C:H02	M00023023B:A05	M00021696C:E02
M00007064B:E09	M00027507C:C06	M00022716A:C01	M00021698A:H03
M00007121A:G04	M00027529B:B11	M00022725D:G05	M00021864C:C07
M00007107C:D02	M00027438D:A03	M00022817D:B09	M00021958A:A04
M00007178D:A10	M00027388A:G05	M00022848D:H09	M00021949D:A05
M00007156D:E11	M00027396C:B06	M00022884D:A07	M00021951B:A01
M00007172D:H03	M00027551C:B07	M00022983A:H04	M00022001B:H10
M00007175D:G02	M00027518B:B07	M00023034B:B10	M00022001D:E06
M00007121D:A11	M00027528A:G03	M00023038D:D04	M00022071D:C08
M00007101C:H01	M00027759B:E11	M00022743C:G05	M00022078B:B04
M00007104D:D10	M00027728A:B03	M00022734C:A03	M00022113B:A12
M00007116A:C08	M00027484A:G03	M00022737D:B02	M00022138C:B07
M00007152A:A10	M00027752B:E05	M00022801A:G04	M00022152A:G05

M00007179B:H04		M00022838B:E05	M00022158C:C08
M00007157B:B04		M00022856A:B09	M00022192B:H07
M00007167C:B10		M00022902C:F11	M00022233C:D11
M00007175B:B11		M00022893D:C06	M00022252A:C01
M00007177B:C02		M00022922D:G06	M00022370A:G07
M00007141A:G08		M00022986B:C02	M00022300A:A05
M00007196D:D02		M00023002D:C12	M00022386D:C04
M00007145C:B05		M00023096C:A03	M00022072D:E12
M00007126D:H01		M00023097A:C03	M00022102D:A10
M00007140C:G12		M00022743C:G06	M00022207C:C01
M00007200A:B12		M00022736B:B03	M00022249C:G09
M00007203C:E06		M00022737B:F12	M00022383C:F05
		M00022831C:F11	M00022384B:E06
		M00022836C:A07	M00022067A:B03
		M00022854D:C04	M00022056B:G12
		M00022860A:A07	M00022084B:C03
		M00022861C:B04	M00022087D:F12
		M00023096A:F03	
		M00023096D:B11	
		M00023097C:D10	

Table 82

ES59	ES60	ES61	ES62
M00001418A:A02	M00001477A:G02	M00004450A:G07	M00005515B:B08
M00003877C:A08	M00003853C:A09	M00004353D:C06	M00005385B:A10
M00003977C:D01	M00001694B:H12	M00004406A:H12	M00005516D:F12
M00004295A:C02	M00001664D:E02	M00004048C:C02	M00005822D:C05
M00001383C:C04	M00003847B:H01	M00004170B:G04	M00004841C:H03
M00001500A:A02	M00001631D:G08	M00004108C:D07	M00005810B:G02
M00003880B:D03	M00004498D:F02	M00004125B:A02	M00007107A:H08
M00003803B:G12	M00001563A:F04	M00004109A:B07	M00004825A:G12
M00003819D:B02	M00001558D:E02	M00004123B:G05	M00005327C:G08
M00004178B:F07	M00004278C:H11	M00004152A:F03	M00005390C:E05

ES63	ES64	ES65	ES66

M00005520A:H11	M00006790D:F10	M00027175D:A05	M00026949A:F04
M00006814D:D09	M00006627C:C02	M00026910C:C05	M00023432D:F09
M00006918D:G08	M00027462D:A12	M00027280D:H01	M00027178B:E04
M00007197D:D12	M00026972A:F04	M00023289D:E06	M00027225B:D03
M00005497C:G08	M00027592D:C05	M00023373A:D01	M00023340B:B07
M00007109D:G01	M00026945B:C10	M00027231A:D01	M00027283C:H12
M00005377C:F07	M00027231C:D08	M00023321A:F07	M00027085C:H12
M00006813B:E04	M00027083D:F06	M00027266C:G12	M00027234C:B05
M00005825A:A10	M00027142A:C01	M00023398D:F10	M00023390A:C04
M00005416B:A01	M00027607A:A09	M00027603C:E02	M00026810A:H04

ES67	**ES68**	**ES69**	**ES70**
M00023340B:H12	M00027642C:D11	M00022714B:D04	M00022709A:C01
M00027237C:D04	M00027202B:B09	M00022838A:H05	M00022413B:D07
M00026809C:D10	M00027459A:G12	M00022392C:H06	M00022467C:H07
M00027386D:C02	M00027250A:C04	M00022363C:D03	M00022561B:B09
M00027343B:H05	M00027499B:G02	M00022205A:C02	M00022214C:E09
M00027356A:H02	M00027053C:B06	M00022717C:F05	M00022697A:C08
M00027363D:A08	M00027598C:D06	M00008015B:D08	M00022682A:F10
M00027364D:E08	M00006989C:B01	M00021625B:G07	M00021841A:E11
M00027618A:B08	M00006837B:H12	M00008100D:C08	M00021691B:E04
M00027628D:D08	M00007202A:A09	M00022669D:G07	M00022477C:C07

ES71	**ES72**	**ES73**	**ES74**
M00022134D:D12	M00008028D:B01	M00022513C:E10	M00023363C:A04
M00022705B:F08	M00021931B:F04	M00022518C:C0 4	M00001401B:A02
M00022903D:H02	M00008097C:E04	M00022544C:D0 8	M00008023C:A06
M00022915C:C09	M00008082B:H10	M00022785C:B1 0	M00022077D:A12
M00007965C:B02	M00008006A:H02	M00022525C:E09	M00023284B:G06
M00022368C:C11	M00022167B:H02	M00022641D:F0 8	M00023369D:C05
M00007937C:E08	M00022509D:A12	M00022923A:A0 9	M00023413D:F04
M00021852C:D12	M00022169A:E11		M00026905A:G11

M00008000D:G11	M00022184D:H07		M00027169D:H06
M00021908B:F03	M00022441B:A06		M00005434D:H02

The deposits described herein are provided merely as convenience to those of skill in the art, and is not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained within the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with the written description of sequences herein. A license may be required to make, use, or sell the deposited material, and no such license is granted hereby.

Retrieval of Individual Clones from Deposit of Pooled Clones. Where the ATCC deposit is composed of a pool of cDNA clones or a library of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones in the pool or library were then deposited as a pool of equal mixtures in the composite deposit. Particular clones can be obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a T_m of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C). Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

EXAMPLE 55

SOURCE OF BIOLOGICAL MATERIALS AND OVERVIEW OF NOVEL POLYNUCLEOTIDES EXPRESSED BY THE BIOLOGICAL MATERIALS

Cell lines and human normal and tumor tissue were used to construct cDNA libraries from mRNA isolated from the cells and tissues. Most sequences were about 275-300 nucleotides in length. The cells lines include Km12L4-A cell line, a high metastatic colon cancer cell line (Morika, W. A. K. et al., *Cancer Research* (1988) *48*:6863). The KM12L4-A cell line is derived from the KM12C cell line. The KM12C cell line, which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B2 surgical specimen (Morikawa et al. *Cancer Res.* (1988) *48*:6863). The KML4-A is a highly metastatic subline derived from KM12C (Yeatman et al. *Nucl. Acids. Res.* (1995) *23*:4007; Bao-Ling et al. *Proc. Annu. Meet. Am. Assoc. Cancer. Res.* (1995) *21*:3269). The KM12C and KM12C-derived cell lines (*e.g.*, KM12L4, KM12L4-A, *etc.*) are well-recognized in the art as model cell lines for the study of colon cancer (see, *e.g.*, Moriakawa et al., *supra*; Radinsky et al. *Clin. Cancer Res.* (1995) 1:19; Yeatman et al., (1995) *supra*; Yeatman et al., *Clin. Exp. Metastasis* (1996) 14:246). These and other cell lines and tissue are described in Table 88.

The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the XBLAST masking program (Claverie "Effective Large-Scale Sequence Similarity Searches," In: Computer Methods for Macromolecular Sequence Analysis, Doolittle, ed., *Meth. Enzymol. 266*:212-227 Academic Press, NY, NY (1996); see particularly Claverie, in "Automated DNA Sequencing and Analysis Techniques" Adams et al., eds., Chap. 36, p. 267 Academic Press, San Diego, 1994 and Claverie et al. *Comput. Chem.* (1993) *17*:191). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, *e.g.*, Alu repeats. The sequences remaining after masking were then used in a BLASTN vs. Genbank search; sequences that exhibited greater than 70% overlap, 99% identity, and a p value of less than 1×10^{-40} were discarded. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a BLASTX vs. NRP (non-redundant proteins) database

search: (1) unknown (no hits in the Genbank search), (2) weak similarity (greater than 45% identity and p value of less than 1 x 10^{-5}), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1 x 10^{-5}). Sequences having greater than 70% overlap, greater than 99% identity, and p value of less than 1 x 10^{-40} were discarded.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a BLAST vs. EST database search was performed and sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1 x 10^{-40} were discarded. Sequences with a p value of less than 1 x 10^{-65} when compared to a database sequence of human origin were also excluded. Second, a BLASTN vs. Patent GeneSeq database was performed and sequences having greater than 99% identity, p value less than 1 x 10^{-40}, and greater than 99% overlap were discarded.

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1 x 10^{-111} in relation to a database sequence of human origin were specifically excluded. The final result provided the 3351 sequences listed in the accompanying Sequence Listing. Each identified polynucleotide represents sequence from at least a partial mRNA transcript. Polynucleotides that were determined to be novel were assigned a sequence identification number.

The novel polynucleotides were assigned sequence identification numbers SEQ ID NOs:9920-12191. The DNA sequences corresponding to the novel polynucleotides are provided in the Sequence Listing.Tables 83 and 84 and 2 provide: 1) the SEQ ID NO assigned to each sequence for use in the present specification or a corresponding number; 2) the sequence name used as an internal identifier of the sequence; 3) the name assigned to the clone from which the sequence was isolated; and 4) the number of the cluster to which the sequence is assigned (Cluster ID; where the cluster ID is 0, the sequence was not assigned to any cluster).

Because the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides of the invention may represent different regions of the same mRNA transcript and the same gene. Thus, if two or more SEQ ID NOs: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene.

EXAMPLE 56

RESULTS OF PUBLIC DATABASE SEARCH TO IDENTIFY FUNCTION OF GENE PRODUCTS

SEQ ID NOs:9920-13270 were translated in all three reading frames to determine the best alignment with the individual sequences. These amino acid sequences and nucleotide sequences are referred to, generally, as query sequences, which are aligned with the individual sequences. Query and individual sequences were aligned using the BLAST programs, available over the world wide web at http://www.ncbi.nlm.nih.gov/BLAST/. Again the sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the XBLAST program for masking low complexity as described above.

Tables 85 and 86 (inserted before the claims) show the results of the alignments. Tables 85 and 86 refer to each sequence by its SEQ ID NO or a corresponding number, the accession numbers and descriptions of nearest neighbors from the Genbank and Non-Redundant Protein searches, and the p values of the search results.

The activity of the polypeptide encoded by SEQ ID NOs:9920-13270 is the same or similar to the nearest neighbor reported in Table 85 or 86. The accession number of the nearest neighbor is reported, providing a reference to the activities exhibited by the nearest neighbor. The search program and database used for the alignment also are indicated as well as a calculation of the p value.

Full length sequences or fragments of the polynucleotide sequences of the nearest neighbors can be used as probes and primers to identify and isolate the full length sequence of SEQ ID NOs: 9920-13270. The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences of SEQ ID NOs: 9920-13270 1.

EXAMPLE 57

MEMBERS OF PROTEIN FAMILIES

The sequences were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein families (and thus represent new members of these protein families) and/or comprising a known functional domain (Table 87). "Start" and "stop" in Table 3 indicate the position within the individual sequences that align with the query sequence having the indicated SEQ ID NO. The direction indicates the orientation of the query sequence with respect to the individual

sequence, where forward (for) indicates that the alignment is in the same direction (left to right) as the sequence provided in the Sequence Listing and reverse (rev) indicates that the alignment is with a sequence complementary to the sequence provided in the Sequence Listing.

Some polynucleotides exhibited multiple profile hits because, for example, the particular sequence contains overlapping profile regions, and/or the sequence contains two different functional domains. These profile hits are described in more detail below.

Ank Repeats (ANK). Some SEQ ID NOs represent polynucleotides encoding an Ank repeat-containing protein. The ankyrin motif is a 33 amino acid sequence named for the protein ankyrin which has 24 tandem 33-amino-acid motifs. Ank repeats were originally identified in the cell-cycle-control protein cdc10 (Breeden et al., *Nature* (1987) *329*:651). Proteins containing ankyrin repeats include ankyrin, myotropin, I-kappaB proteins, cell cycle protein cdc10, the Notch receptor (Matsuno et al., *Development* (1997) *124*(*21*):4265); G9a (or BAT8) of the class III region of the major histocompatibility complex (*Biochem J. 290*:811-818, 1993), FABP, GABP, 53BP2, Lin12, glp-1, SW14, and SW16. The functions of the ankyrin repeats are compatible with a role in protein-protein interactions (Bork, *Proteins* (1993) *17*(*4*):363; Lambert and Bennet, *Eur. J. Biochem.* (1993) *211*:1; Kerr et al., *Current Op. Cell Biol.* (1992) *4*:496; Bennet et al., *J. Biol. Chem.* (1980) *255*:6424).

ATPases Associated with Various Cellular Activities (ATPases). Some SEQ ID NOs correspond to a sequence that encodes a novel member of the "ATPases Associated with diverse cellular Activities" (AAA) protein family. The AAA protein family is composed of a large number of ATPases that share a conserved region of about 220 amino acids that contains an ATP-binding site (Froehlich et al., *J. Cell Biol.* (1991) *114*:443; Erdmann et al., *Cell* (1991) *64*:499; Peters et al., *EMBO J.* (1990) *9*:1757; Kunau et al., *Biochimie* (1993) *75*:209-224; Confalonieri et al., *BioEssays* (1995) *17*:639; http://yeamob.pci.chemie.uni-tuebingen.de/AAA/Description.html). The proteins that belong to this family either contain one or two AAA domains. In general, the AAA domains in these proteins act as ATP-dependent protein clamps (Confalonieri et al. (1995) *BioEssays 17*:639). In addition to the ATP-binding 'A' and 'B' motifs, which are located in the N-terminal half of this domain, there is a highly conserved region located in the central part of the domain which was used in the development of the signature pattern.

Bromodomain (bromodomain). One SEQ ID NOrepresents a polynucleotide encoding a polypeptide having a bromodomain region (Haynes et al., 1992, Nucleic Acids Res. 20:2693-2603, Tamkun et al., 1992, Cell 68:561-572, and Tamkun, 1995, Curr. Opin.

Genet. Dev. 5:473-477), which is a conserved region of about 70 amino acids. The bromodomain is thought to be involved in protein-protein interactions and may be important for the assembly or activity of multicomponent complexes involved in transcriptional activation.

Basic Region Plus Leucine Zipper Transcription Factors (BZIP). Some SEQ ID NOs represent polynucleotides encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization.

EF Hand (EFhand). Some SEQ ID NOs correspond to polynucleotides encoding a novel protein in the family of EF-hand proteins. Many calcium-binding proteins belong to the same evolutionary family and share a type of calcium-binding domain known as the EF-hand (Kawasaki et al., *Protein. Prof.* (1995) *2*:305-490). This type of domain consists of a twelve residue loop flanked on both sides by a twelve residue alpha-helical domain. In an EF-hand loop the calcium ion is coordinated in a pentagonal bipyramidal configuration. The six residues involved in the binding are in positions 1, 3, 5, 7, 9 and 12; these residues are denoted by X, Y, Z, -Y, -X and -Z. The invariant Glu or Asp at position 12 provides two oxygens for liganding Ca (bidentate ligand).

Ets Domain (Ets_Nterm). One SEQ ID NO represents a polynucleotide encoding a polypeptide with N-terminal homology in ETS domain. Proteins of this family contain a conserved domain, the "ETS-domain," that is involved in DNA binding. The domain appears to recognize purine-rich sequences; it is about 85 to 90 amino acids in length, and is rich in aromatic and positively charged residues (Wasylyk, et al., *Eur. J. Biochem.* (1993) *211*:718). The *ets* gene family encodes a novel class of DNA-binding proteins, each of which binds a specific DNA sequence and comprises an *ets* domain that specifically interacts with sequences containing the common core tri-nucleotide sequence GGA. In addition to an *ets* domain, native *ets* proteins comprise other sequences which can modulate the biological specificity of the protein. *Ets* genes and proteins are involved in a variety of essential biological processes including cell growth, differentiation and development, and three members are implicated in oncogenic process.

G-Protein Alpha Subunit (G-alpha). One SEQ ID NO represents a polynucleotide encoding a novel polypeptide of the G-protein alpha subunit family. Guanine nucleotide binding proteins (G-proteins) are a family of membrane-associated proteins that couple extracellularly-activated integral-membrane receptors to intracellular effectors, such as ion

channels and enzymes that vary the concentration of second messenger molecules. G-proteins are composed of 3 subunits (alpha, beta and gamma) which, in the resting state, associate as a trimer at the inner face of the plasma membrane. The alpha subunit binds GTP and exhibits GTPase activity. G-protein alpha subunits are 350-400 amino acids in length and have molecular weights in the range 40-45 kDa. Seventeen distinct types of alpha subunit have been identified in mammals, and fall into 4 main groups on the basis of both sequence similarity and function: alpha-s, alpha-q, alpha-i and alpha-12 (Simon et al., *Science* (1993) *252*:802). They are often N-terminally acylated, usually with myristate and/or palmitoylate, and these fatty acid modifications can be important for membrane association and high- affinity interactions with other proteins.

Helicases conserved C-terminal domain (helicase_C). Some SEQ ID NOs represent polynucleotides encoding novel members of the DEAD/H helicase family. A number of eukaryotic and prokaryotic proteins have been characterized (Schmid S.R., et al., *Mol. Microbiol.* (1992) *6*:283; Linder P., et al., *Nature* (1989) *337*:121; Wassarman D.A., et al., *Nature* (1991) *349*:463) on the basis of their structural similarity. All are involved in ATP-dependent, nucleic-acid unwinding. All DEAD box family members of the above proteins share a number of conserved sequence motifs, some of which are specific to the DEAD family while others are shared by other ATP-binding proteins or by proteins belonging to the helicases 'superfamily' (Hodgman T.C., *Nature* (1988) *333*:22 and *Nature* (1988) *333*:578 (Errata). One of these motifs, called the "D-E-A-D-box", represents a special version of the B motif of ATP-binding proteins. Some other proteins belong to a subfamily which have His instead of the second Asp and are thus said to be "D-E-A-H-box" proteins (Wassarman D.A., et al., *Nature* (1991) *349*:463; Harosh I., et al., *Nucleic Acids Res.* (1991) *19*:6331; Koonin E.V. et al., *J. Gen. Virol.* (1992) *73*:989.

Homeobox domain (homeobox). Some SEQ ID NOs represent polynucleotides encoding proteins having a homeobox domain. The homeobox is a protein domain of 60 amino acids (Gehring In: Guidebook to the Homeobox Genes, Duboule D., Ed., pp. 1-10, Oxford University Press, Oxford, (1994); Buerglin In: Guidebook to the Homeobox Genes, pp25-72, Oxford University Press, Oxford, (1994); Gehring, *Trends Biochem. Sci.* (1992) 17:277-280; Gehring et al., *Annu. Rev. Genet.* (1986) *20*:147-173; Schofield, *Trends Neurosci.* (1987) *10*:3-6) first identified in a number of Drosophila homeotic and segmentation proteins. It is extremely well conserved in many other animals, including vertebrates. This domain binds DNA through a helix-turn-helix type of structure. Several proteins that contain a homeobox domain play an important role in development. Most of these proteins are sequence-specific DNA-binding transcription factors. The homeobox

domain is also very similar to a region of the yeast mating type proteins. These are sequence-specific DNA-binding proteins that act as master switches in yeast differentiation by controlling gene expression in a cell type-specific fashion.

A schematic representation of the homeobox domain is shown below. The helix-turn-helix region is shown by the symbols 'H' (for helix), and 't' (for turn).

```
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxHHHHHHHHHtttHHHHHHHHHHxxxxxxxxxxx
1                                                                60
```

The pattern detects homeobox sequences 24 residues long and spans positions 34 to 57 of the homeobox domain.

MAP kinase kinase (mkk). Some SEQ ID NOs represent novel members of the MAP kinase kinase family. MAP kinases (MAPK) are involved in signal transduction, and are important in cell cycle and cell growth controls. The MAP kinase kinases (MAPKK) are dual-specificity protein kinases which phosphorylate and activate MAP kinases. MAPKK homologues have been found in yeast, invertebrates, amphibians, and mammals. Moreover, the MAPKK/MAPK phosphorylation switch constitutes a basic module activated in distinct pathways in yeast and in vertebrates. MAPKKs are essential transducers through which signals must pass before reaching the nucleus. For review, see, *e.g.*, Biologique *Biol Cell* (1993) *79*:193-207; Nishida et al., *Trends Biochem Sci* (1993) *18*:128-31; Ruderman, *Curr Opin Cell Biol* (1993) *5*:207-13; Dhanasekaran et al., *Oncogene* (1998) *17*:1447-55; Kiefer et al., *Biochem Soc Trans* (1997) *25*:491-8; and Hill, *Cell Signal* (1996) *8*:533-44.

Protein Kinase (protkinase). Some SEQ ID NOs represent polynucleotides encoding protein kinases. Protein kinases catalyze phosphorylation of proteins in a variety of pathways, and are implicated in cancer. Eukaryotic protein kinases (Hanks S.K., et al., *FASEB J.* (1995) 9:576; Hunter T., *Meth. Enzymol.* (1991) *200*:3; Hanks S.K., et al., *Meth. Enzymol.* (1991) *200*:38; Hanks S.K., *Curr. Opin. Struct. Biol.* (1991) *1*:369; Hanks S.K. et al., *Science* (1988) *241*:42) are enzymes that belong to a very extensive family of proteins which share a conserved catalytic core common to both serine/threonine and tyrosine protein kinases. There are a number of conserved regions in the catalytic domain of protein kinases. The first region, which is located in the N-terminal extremity of the catalytic domain, is a glycine-rich stretch of residues in the vicinity of a lysine residue, which has been shown to be involved in ATP binding. The second region, which is located in the central part of the catalytic domain, contains a conserved aspartic acid residue which is important for the catalytic activity of the enzyme (Knighton D.R. et al., *Science* (1991)

253:407). The protein kinase profile includes two signature patterns for this second region: one specific for serine/threonine kinases and the other for tyrosine kinases. A third profile is based on the alignment in (Hanks S.K. et al., *FASEB J.* (1995) *9*:576) and covers the entire catalytic domain.

If a protein analyzed includes two of the above protein kinase signatures, the probability of it being a protein kinase is close to 100%.

Ras family proteins (ras). Some SEQ ID NOs represent polynucleotides encoding novel members of the ras family of small GTP/GDP-binding proteins (Valencia et al., 1991, Biochemistry 30:4637-4648). Ras family members generally require a specific guanine nucleotide exchange factor (GEF) and a specific GTPase activating protein (GAP) as stimulators of overall GTPase activity. Among ras-related proteins, the highest degree of sequence conservation is found in four regions that are directly involved in guanine nucleotide binding. The first two constitute most of the phosphate and Mg2+ binding site (PM site) and are located in the first half of the G-domain. The other two regions are involved in guanosine binding and are located in the C-terminal half of the molecule. Motifs and conserved structural features of the ras-related proteins are described in Valencia et al., 1991, Biochemistry 30:4637-4648.

Thioredoxin family active site (Thioredox). One SEQ ID NO represents a polynucleotide encoding a protein having a thioredoxin family active site. Thioredoxins (Holmgren A., *Annu. Rev. Biochem.* (1985) *54*:237; Gleason F.K. et al., *FEMS Microbiol. Rev.* (1988) *54*:271; Holmgren, *A. J. Biol. Chem.* (1989) *264*:13963; Eklund H. et al., *Proteins* (1991) *11*:13) are small proteins of approximately one hundred amino- acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is well conserved.

Trypsin (trypsin). One SEQ ID NO corresponds to a novel serine protease of the trypsin family. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature* (1988) *334*:528).

<u>WD Domain, G-Beta Repeats (WD_domain)</u>. Some SEQ ID NOs represent novel members of the WD domain/G-beta repeat family. Beta-transducin (G-beta) is one of the three subunits (alpha, beta, and gamma) of the guanine nucleotide-binding proteins (G proteins) which act as intermediaries in the transduction of signals generated by transmembrane receptors (Gilman, *Annu. Rev. Biochem.* (1987) *56*:615). The alpha subunit binds to and hydrolyzes GTP; the functions of the beta and gamma subunits are less clear but they seem to be required for the replacement of GDP by GTP as well as for membrane anchoring and receptor recognition. In higher eukaryotes, G-beta exists as a small multigene family of highly conserved proteins of about 340 amino acid residues. Structurally, G-beta consists of eight tandem repeats of about 40 residues, each containing a central Trp-Asp motif (this type of repeat is sometimes called a WD-40 repeat).

<u>wnt Family of Developmental Signaling Proteins (Wnt_dev_sign)</u>. One SEQ ID NO corresponds to a novel member of the wnt family of developmental signaling proteins. Wnt-1 (previously known as int-1), the seminal member of this family, (Nusse R., *Trends Genet.* (1988) *4*:291) is thought to play a role in intercellular communication and seems to be a signalling molecule important in the development of the central nervous system (CNS). All wnt family proteins share the following features characteristics of secretory proteins: a signal peptide, several potential N-glycosylation sites and 22 conserved cysteines that are probably involved in disulfide bonds. The Wnt proteins seem to adhere to the plasma membrane of the secreting cells and are therefore likely to signal over only few cell diameters.

<u>Protein Tyrosine Phosphatase (Y_phosphatase)</u>. One SEQ ID NO represents a polynucleotide encoding a protein tyrosine kinase. Tyrosine specific protein phosphatases (EC 3.1.3.48) (PTPase) (Fischer et al., *Science* (1991) *253*:401; Charbonneau et al., *Annu. Rev. Cell Biol.* (1992) *8*:463; Trowbridge, *J. Biol. Chem.* (1991) *266*:23517; Tonks et al., *Trends Biochem. Sci.* (1989) *14*:497; and Hunter, *Cell* (1989) *58*:1013) catalyze the removal of a phosphate group attached to a tyrosine residue. These enzymes are very important in the control of cell growth, proliferation, differentiation and transformation. Multiple forms of PTPase have been characterized and can be classified into two categories: soluble PTPases and transmembrane receptor proteins that contain PTPase domain(s). Structurally, all known receptor PTPases are made up of a variable length extracellular domain, followed by a transmembrane region and a C-terminal catalytic cytoplasmic domain. PTPase domains consist of about 300 amino acids. The search of two conserved cysteines has been shown to be absolutely required for activity. Furthermore, a

number of conserved residues in its immediate vicinity have also been shown to be important.

Zinc Finger, C2H2 Type (Zincfing_C2H2). Some SEQ ID NOs correspond to polynucleotides encoding novel members of the of the C2H2 type zinc finger protein family. Zinc finger domains (Klug et al., *Trends Biochem. Sci.* (1987) *12*:464; Evans et al., *Cell* (1988) *52*:1; Payre et al., *FEBS Lett.* (1988) *234*:245; Miller et al., *EMBO J.* (1985) *4*:1609; and Berg, *Proc. Natl. Acad. Sci.* USA (1988) *85*:99) are nucleic acid-binding protein structures. In addition to the conserved zinc ligand residues, it has been shown that a number of other positions are also important for the structural integrity of the C2H2 zinc fingers. (Rosenfeld et al., *J. Biomol. Struct. Dyn.* (1993) *11*:557) The best conserved position is found four residues after the second cysteine; it is generally an aromatic or aliphatic residue.

Src homology 2. Some SEQ ID NOs represent polynucleotides encoding novel members of the family of Src homology 2 (SH2) proteins. The Src homology 2 (SH2) domain is a protein domain of about 100 amino acid residues first identified as a conserved sequence region between the oncoproteins Src and Fps (Sadowski I. et al., *Mol. Cell. Biol.* *6*:4396-4408 (1986)). Similar sequences are found in many other intracellular signal-transducing proteins (Russel R.B. et al., *FEBS Lett.* *304*:15-20 (1992)). SH2 domains function as regulatory modules of intracellular signalling cascades by interacting with high affinity to phosphotyrosine-containing target peptides in a sequence-specific and phosphorylation-dependent manner (Marangere L.E.M., Pawson T., *J. Cell Sci. Suppl.* *18*:97-104 (1994); Pawson T., Schlessinger J., *Curr. Biol.* *3*:434-442 (1993); Mayer B.J., Baltimore D., *Trends Cell. Biol.* *3*:8-13 (1993); Pawson T., *Nature 373*:573-580 (1995)).

The SH2 domain has a conserved 3D structure consisting of two alpha helices and six to seven beta-strands. The core of the domain is formed by a continuous beta-meander composed of two connected beta-sheets (Kuriyan J., Cowburn D., *Curr. Opin. Struct. Biol.* *3*:828-837(1993)). The profile to detect SH2 domains is based on a structural alignment consisting of 8 gap-free blocks and 7 linker regions totaling 92 match positions.

Src homology 3. Some SEQ ID NOs represent polynucleotides encoding novel members of the family of Src homology 3 (SH3) proteins. The Src homology 3 (SH3) domain is a small protein domain of about 60 amino acid residues first identified as a conserved sequence in the non-catalytic part of several cytoplasmic protein tyrosine kinases (*e.g.*, Src, Abl, Lck) (Mayer B.J. et al., *Nature 332*:272-275 (1988)). Since then, it has been found in a great variety of other intracellular or membrane-associated proteins (Musacchio A. et al., *FEBS Lett.* *307*:55-61 (1992); Pawson T., Schlessinger J., Curr. Biol.

3:434-442 (1993); Mayer B.J., Baltimore D., *Trends Cell Biol. 3*:8-13 (1993); Pawson T., *Nature 373*:573-580 (1995)).

The SH3 domain has a characteristic fold which consists of five or six beta strands arranged as two tightly packed anti-parallel beta sheets. The linker regions may contain short helices (Kuriyan J., Cowburn D., *Curr. Opin. Struct. Biol. 3*:828-837 (1993)).

The function of the SH3 domain may be to mediate assembly of specific protein complexes via binding to proline-rich peptides (Morton C.J., Campbell I.D., *Curr. Biol. 4*:615-617 (1994)).

In general SH3 domains are found as single copies in a given protein, but there are a significant number of proteins with two SH3 domains and a few with 3 or 4 copies.

Fibronectin type III. Some SEQ ID NOs represent polynucleotides encoding novel members of the family of fibronectin type III proteins. A number of receptors for lymphokines, hematopoeitic growth factors and growth hormone-related molecules have been found to share a common binding domain. (Bazan J.F., *Biochem. Biophys. Res. Commun. 164*:788-795 (1989); Bazan J.F., *Proc. Natl. Acad. Sci. U.S.A. 87*:6934-6938 (1990); Cosman D. et al., *Trends Biochem. Sci. 15*:265-270 (1990); d'Andrea A.D., Fasman G.D., Lodish H.F., *Cell 58*:1023-1024 (1989); d'Andrea A.D., Fasman G.D., Lodish H.F., *Curr. Opin. Cell Biol. 2*:648-651 (1990)).

The conserved region constitutes all or part of the extracellular ligand- binding region and is about 200 amino acid residues long. In the N-terminal of this domain there are two pairs of cysteines known, in the growth hormone receptor, to be involved in disulfide bonds.

```
+-----------------------------------xxxxxxx-------------------------+
|C C         C C       Extracellular     XXXXXXX      Cytoplasmic  | +
-|-|-------|--|------------------------xxxxxxx--------------------------+
 | |        | |                     Transmembrane
+-+        +--+
```

Two patterns detect this family of receptors. The first one is derived from the first N-terminal disulfide loop, the second is a tryptophan-rich pattern located at the C-terminal extremity of the extracellular region.

LIM domain containing proteins. Some SEQ ID NOs represent polynucleotides encoding novel members of the family of LIM domain containing proteins. A number of proteins contain a conserved cysteine-rich domain of about 60 amino-acid residues. (Freyd

G. et al., *Nature 344*:876-879 (1990); Baltz R. et al., *Plant Cell 4*:1465-1466 (1992); Sanchez-Garcia I., Rabbitts T.H., *Trends Genet. 10*:315-320 (1994)).

In the LIM domain, there are seven conserved cysteine residues and a histidine.

C2 domain (protein kinase C like). Some SEQ ID NOs represent polynucleotides encoding novel members of the family of C2 domain containing proteins. Some isozymes of protein kinase C (PKC) contain a domain, known as C2, of about 116 amino-acid residues, which is located between the two copies of the C1 domain (that bind phorbol esters and diacylglycerol) and the protein kinase catalytic domain. (Azzi A. et al., *Eur. J. Biochem. 208*:547-557 (1992); Stabel S., Semin. Cancer Biol. 5:277-284 (1994)).

The C2 domain is involved in calcium-dependent phospholipid binding (Davletov B.A., Suedhof T.C., *J. Biol. Chem. 268*:26386-26390 (1993)). Since domains related to the C2 domain are also found in proteins that do not bind calcium, other putative functions for the C2 domain include binding to inositol-1,3,5-tetraphosphate. (Fukuda M., et al., *J. Biol. Chem. 269*:29206-29211 (1994).)

The consensus pattern for the C2 domain is located in a conserved part of that domain, the connecting loop between beta strands 2 and 3. The profile for the C2 domain covers the total domain.

Serine proteases, trypsin family, active sites. One SEQ ID NO represents a polynucleotide encoding a novel member of the family of serine protease, trypsin proteins. The catalytic activity of the serine proteases from the trypsin family is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which itself is hydrogen-bonded to a serine. The sequences in the vicinity of the active site serine and histidine residues are well conserved in this family of proteases (Brenner S., *Nature 334*:528-530 (1988)).

RNA Recognition Motif Domain (RRM, RBD, or RNP). Some SEQ ID NOs represent polynucleotides encoding novel members of the family of RNA recognition motif domain proteins (Bandziulis R.J. et al., *Genes Dev. 3*:431-437 (1989); Dreyfuss G. et al., *Trends Biochem. Sci. 13*:86-91 (1988)).

Inside the putative RNA-binding domain there are two regions which are highly conserved. The first one is a hydrophobic segment of six residues (which is called the RNP-2 motif); the second one is an octapeptide motif (which is called RNP-1 or RNP-CS). The position of both motifs in the domain is shown in the following schematic representation:

```
xxxxxxx#####xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx#######xxxxxxxxxxxxxxxxxxxxxxxxxx
        RNP-2                                    RNP-1
```

Phosphatidylinositol-specific phospholipase C, Y Domain. One SEQ ID NO represents a polynucleotide encoding a novel member of the phosphatidylinositol-specific phospholipase C, Y domain family of proteins. Phosphatidylinositol-specific phospholipase C (EC3.1.4.11), a eukaryotic intracellular enzyme, plays an important role in signal transduction processes (Meldrum E. et al., *Biochim. Biophys. Acta 1092*:49-71 (1991)). It catalyzes the hydrolysis of 1-phosphatidyl-D-myo-inositol-3,4,5- triphosphate into the second messenger molecules diacylglycerol and inositol-1,4,5-triphosphate. This catalytic process is tightly regulated by reversible phosphorylation and binding of regulatory proteins (Rhee S.G., Choi K.D., *Adv. Second Messenger Phosphoprotein Res. 26*:35-61 (1992); Rhee S.G., Choi K.D., *J. Biol. Chem. 267*:12393-12396 (1992); Sternweis P.C., Smrcka A.V., Trends Biochem. *Sci. 17*:502-506 (1992)).

All eukaryotic PI-PLCs contain two regions of homology, referred to as "X-box" and "Y-box". The order of these two regions is the same (NH2-X-Y-COOH), but the spacing is variable. In most isoforms, the distance between these two regions is only 50-100 residues but in the gamma isoforms one PH domain, two SH2 domains, and one SH3 domain are inserted between the two PLC-specific domains. The two conserved regions have been shown to be important for the catalytic activity. At the C-terminal of the Y-box, there is a C2 domain possibly involved in Ca-dependent membrane attachment.

Serine Carboxypeptidases. One SEQ ID NO represents a polynucleotide encoding a novel member of the serine carboxypeptidases family of proteins. Carboxypeptidases may be either metallo carboxypeptidases or serine carboxypeptidases (EC 3.4.16.5 and EC 3.4.16.6). The catalytic activity of the serine carboxypeptidases, like that of the trypsin family serine proteases, is provided by a charge relay system involving an aspartic acid residue hydrogen-bonded to a histidine, which is itself hydrogen-bonded to a serine (Liao D.I., Remington S.J., *J. Biol. Chem. 265*:6528-6531 (1990)).

dsrm Double-Stranded RNA Binding Motif. One SEQ ID NO represents a polynucleotide encoding a novel member of the dsrm double-stranded RNA binding motif proteins. In eukaryotic cells, a multitude of RNA-binding proteins play key roles in the posttranscriptional regulation of gene expression. Characterization of these proteins has led to the identification of several RNA-binding motifs. Several human and other vertebrate

genetic disorders are caused by aberrant expression of RNA-binding proteins. (C. G. Burd & G. Dreyfuss, *Science 265*: 615-621 (1994)).

Proteins containing double stranded RNA binding motifs bind to specific RNA targets. Double stranded RNA binding motifs are exemplified by interferon-induced protein kinase in humans, which is part of the cellular response to dsRNA.

Some SEQ ID NOs encode members of the 4 trans-membrane integral membrane protein family. This family consists of type III proteins, which are integral membrane proteins that contain a N-terminal membrane-anchoring domain that is not cleaved during biosynthesis, and which functions as a translocation signal and a membrane anchor. The proteins also have three additional transmembrane regions. .

One SEQ ID NO encodes a polypeptide having a calpain large subunit, domain III. Calpains are a family of intracellular proteases that play a variety of biological roles. Calpain 3, also known as p94, is predominantly expressed in skeletal muscle and plays a role in limb-girdle muscular dystrophy type 2A. (Sorimachi, H. et al., Biochem. J. 328:721-732, 1997).

Some SEQ ID NOs encode polypeptides having a C3HC4 type zinc finger domain (RING finger), which is a cysteine-rich domain of 40 to 60 residues that binds two atoms of zinc, and is believed to be involved in mediating protein-protein interactions. Mammalian proteins of this family include V(D)J recombination activating protein, which activates the rearrangement of immunoglobulin and T-cell receptor genes; breast cancer type 1 susceptibility protein (BRCA1); bmi-1 proto-oncogene; cbl proto-oncogene; and mel-18 protein, which is expressed in a variety of tumor cells and is a transcriptional repressor that recognizes and binds a specific DNA sequence.

One SEQ ID NO encodes a eukaryotic transcription factor with a fork head domain, of about 100 amino acid residues. Proteins of this group are transcription factors, including mammalian transcription factors HNF-3-alpha, -beta, and -gamma; interleukin-enhancer binding factor; and HTLF, which binds to a region of human T-cell leukemia virus long terminal repeat.

One SEQ ID NO encodes a polypeptide having a PDZ domain. Several dozen signaling proteins belong to this group of proteins that have 80-100 residue repeats known as PDZ domains. Several of the proteins interact with the C-terminal tetrapeptide motifs X-Ser/Thr/X-Val-COO- of ion channels and/or receptors. (Ponting, C. P., Protein Sci. 6;464-468, 1997.)

One SEQ ID NO encodes a polypeptide in the family of phorbol esters/glycerol binding proteins. Phorbol esters (PE) are analogues of diacylglycerol (DAG) and potent

tumor promoters. DAG activates a family of serine-threonine protein kinases, known as protein kinase C. The N-terminal region of protein kinase C binds PE and DAG, and contains one or two copies of a cysteine-rich domain of about 50 amino acid residues. Other proteins having this domain include diacylglycerol kinase; the vav oncogene; and N-chimaerin, a brain-specific protein. The DAG/PE binding domain binds two zinc ions through the six cysteines and two histidines that are conserved in the domain.

One SEQ ID NO encodes a polypeptide having a WW/rsp5/WWP domain. The protein is named for the presence of conserved aromatic positions, generally tryptophan, as well as a conserved proline. Proteins having the domain include dystrophin, vertebrate YAP protein, and IQGAP, a human GTPase activating protein which acts on ras.

One SEQ ID NO encodes a member of the dual specificity phosphatase family, having a catalytic domain, and some SEQ IDS NOs encode members of the protein tyrosine phosphatase family. These families are related and classified as tyrosine specific protein phosphatases. The enzymes catalyze the removal of a phosphate group from a tyrosine residue, and are important in the control of cell growth, proliferation, differentiation, and transformation.

Table 87

SEQ ID	Start	Stop	Score	Direction	Description
9948	295	421	5872	For	mkk like kinases
9949	31	182	3943	For	Basic region plus leucine zipper transcription factors
9950	298	397	5625	For	mkk like kinases
10105	175	395	7660	For	SH2 Domain
10106	358	432	4320	For	Ank repeat
10115	37	322	6049	For	mkk like kinases
10153	23	121	4607	For	SH3 Domain
10227	110	172	4150	For	Zinc finger, C2H2 type
10329	42	191	4036	For	Basic region plus leucine zipper transcription factors
10350	71	428	5538	Rev	ATPases Associated with Various Cellular Activities
10471	116	288	3930	Rev	Basic region plus leucine zipper transcription factors
10558	157	561	5797	For	ATPases Associated with Various Cellular Activities
10665	209	427	5379	For	Fibronectin type III domain
10687	116	288	3930	For	Basic region plus leucine zipper transcription factors
10726	339	392	3620	For	Zinc finger, C2H2 type
10739	341	406	2930	Rev	EF-hand
10741	108	262	4179	For	Basic region plus leucine zipper transcription factors
10755	158	353	4430	For	Basic region plus leucine zipper transcription factors
11076	41	444	5279	Rev	protein kinase
11111	186	416	5469	For	Fibronectin type III domain
11187	238	315	3540	For	Ank repeat
11188	79	240	11640	For	LIM domain containing proteins
11207	73	234	3953	For	Basic region plus leucine zipper transcription factors
11228	248	404	8226	for	LIM domain containing proteins